

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 organises acoustic coordinates, temporal requests and named calculations so that their inputs and dependencies can be inspected and revised. 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 a compact acoustic baseline and cross-corpus arousal associations. Two controlled studies extend this evidence beyond musical excerpts. Across 1,146 renderings of one black-bird phrase, relative flute–piano displacement reversed between bird and transcription references; terminal mappings and averaging concealed some underlying changes. Across 342 synthetic speech and song recordings, multichannel responses retained instruction information that native affect scores incompletely expressed, with decoder rankings sensitive to regularization. Spectral interventions increased native valence; restoring source consonance features suppressed 94.2–94.8% of the mean increase without correcting the primary instruction ordering. Together, these findings establish an inspectable experimental framework linking response geometry, fitted information and computational dependencies. Cognitive and physiological interpretations require corresponding observations; 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

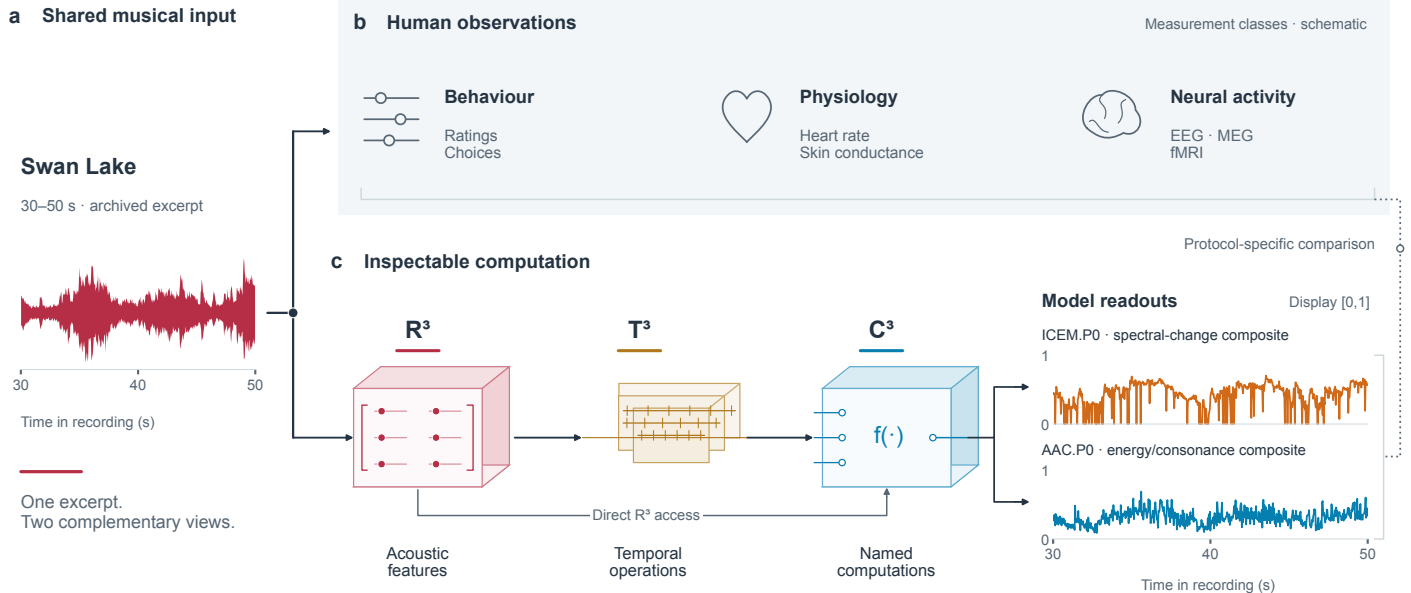


Figure 1. Music as a setting for inspectable computation. The Swan Lake excerpt supplies both paths. Observation symbols denote measurement classes, not collected participant data. R^3 features enter T^3 operations and C^3 calculations, including direct R^3 access. Depth is schematic. Curves retain their 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 reveal reference-dependent sensitivity; and speech experiments separate distributed instruction information from native affect direction. 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.

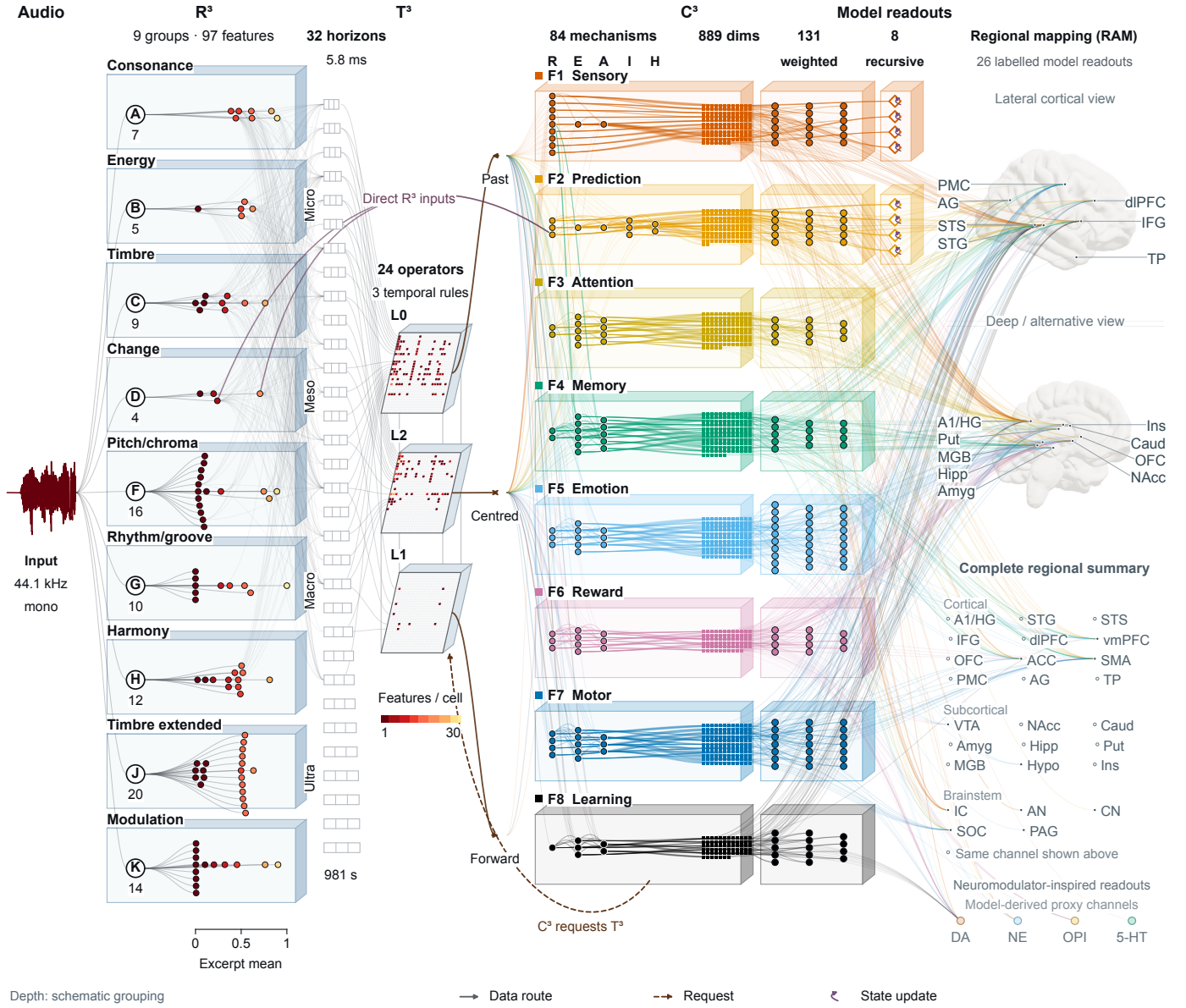


Figure 2. Musical Intelligence architecture. The current configuration contains 84 mechanisms, 889 output dimensions, 131 weighted readouts and eight selected recursive classes. Surfaces group computations; depth is schematic. Solid paths show data routes, dashed arrows requests and loops local state. 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

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

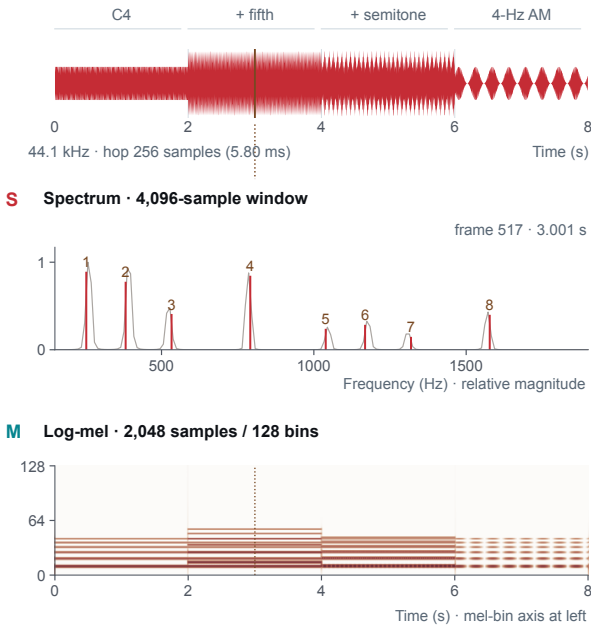
where a is audio, F_θ the engine, O_π the observation mapping and c_π a covariate. When O_π is fitted, prediction depends on engine and readout. Regression estimates the readout weights separately; an observation mapping need not forecast events.

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.

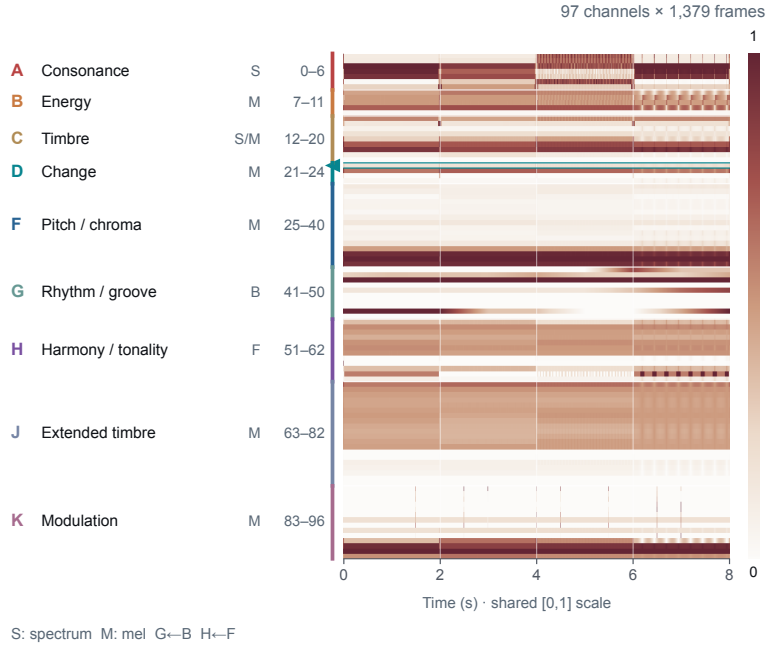
The executor visits each mechanism once, in processing-depth order, retaining discovery order when depths coincide. A mechanism can use only the upstream outputs already available at that point; unavailable inputs invoke its implementation's fallback. A listed connection therefore affects the calculation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

R³ · Acoustic representation

a Two routes from one waveform

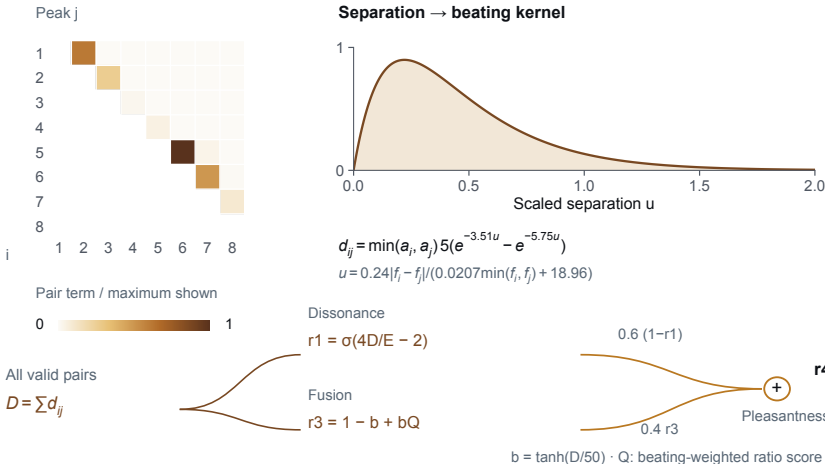


b A field of named coordinates



c Inside the pairwise calculation

Eight displayed peaks → 28 pairs



d Different responses, one clock

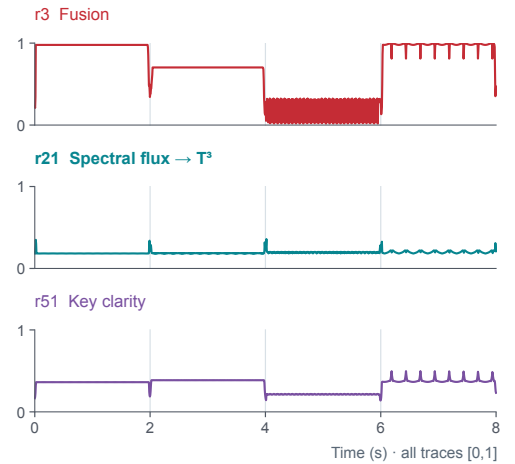


Figure 3. From sound to named acoustic coordinates. (a) The preserved 8-s waveform, saved spectrum at 3.001 s and log-mel field distinguish spectral (S) and mel (M) routes; window lengths differ from the 256-sample hop. (b) All 97 channels and 1,379 frames retain their [0,1] scale. S supplies A and C's 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.

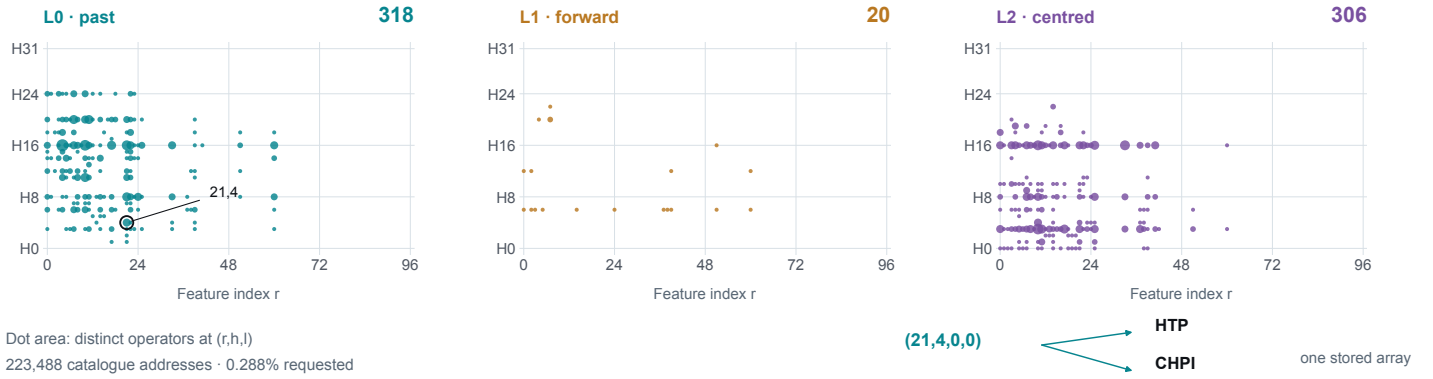
These definitions accompany both the direct inputs to C³ and the series processed by T³.

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.

T³ · Temporal addressing and operations

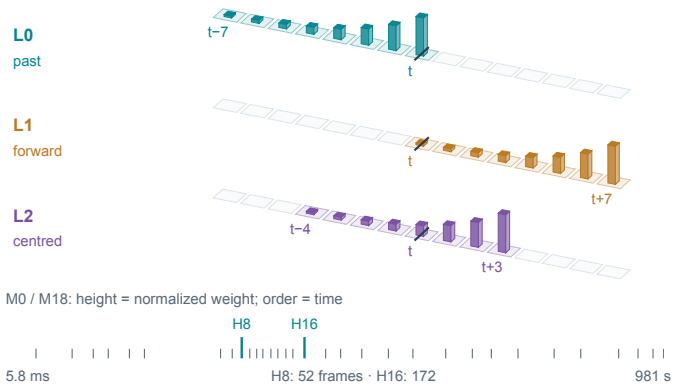
a The instantiated address atlas

Requested here: 34 features · 23 horizons · 19 operators



b The samples behind an address

Eight-sample illustration; this is not H8



c An open operator repertoire

Level / distribution	Change	Structure / events
M0 weighted mean M1 mean M2 sample SD M3 median M4 maximum M5 range M6 skewness M7 kurtosis	M8 velocity M9 mean velocity M10 velocity SD M11 acceleration M12 mean accel. M13 accel. SD M16 curvature M18 trend	M14 periodicity M15 smoothness M17 shape period M19 stability M20 entropy M21 zero crossings M22 peaks M23 symmetry

$M0: \sum_j w_j x_j \quad \sum_j w_j = 1$
 $M2: \text{clip}(\text{sample SD} / 0.25, 0, 1)$
Names retain distinct constants and scaling.

d Hold r21 fixed; change one address field

Query order: (r, h, m, l)



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

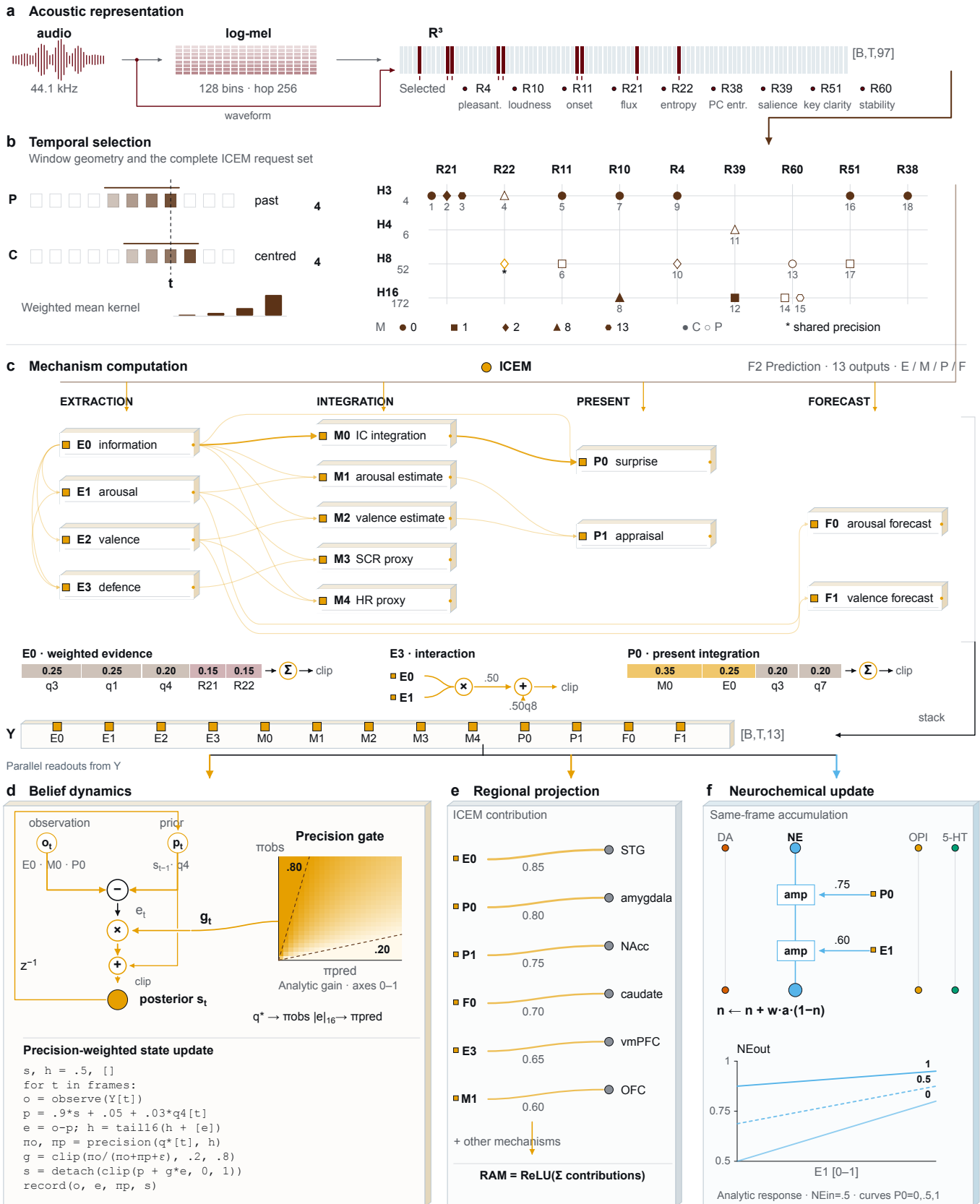


Figure 5. ICEM computation and parallel readouts. (a,b) Nine R³ channels and 18 temporal requests supply (c) 13 named outputs; insets expose weights and interactions. (d–f) Recursive belief, regional accumulation and ordered NE updates form parallel branches. The starred q^* precision input belongs to 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

C³ 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 speech response spectra. These studies distinguish correspondence, represented information and implemented dependencies. 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-adaptor 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

$$|\Delta E_0| \leq 0.25|\Delta q|, \quad |\Delta A_{\text{STG}}| \leq 0.85|\Delta E_0|.$$

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

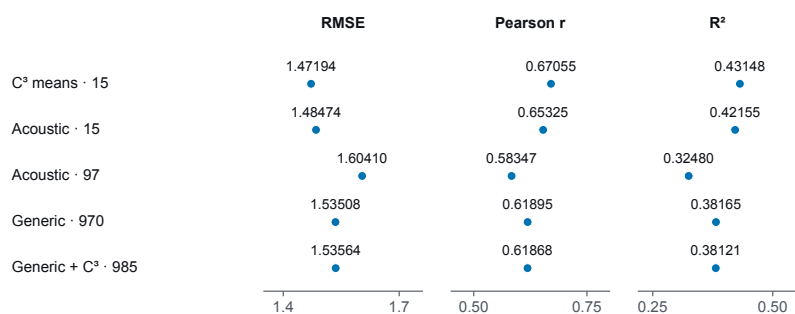
Equation inspection identifies inputs relevant to testing each interpretation. DAPP1 combines inverse roughness, tonalness, sensory pleasantness, inverse spectral entropy and an upstream composite, without developmental-history inputs. MMP2 combines past warmth stability, a rhythmic summary and centred 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

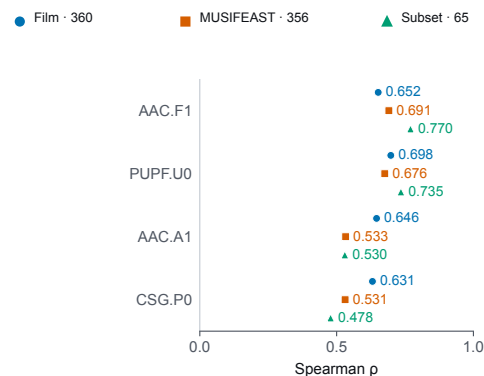
a Grouped tension prediction

110 entries · 40 soundtracks · 5 grouped folds



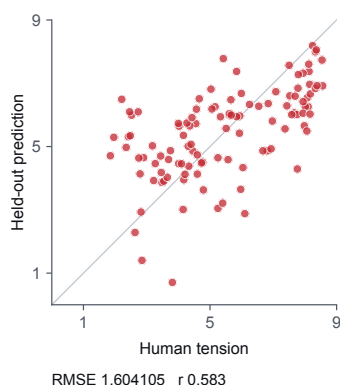
Inputs shown after each name. Lower RMSE is better.

b Arousal across corpora

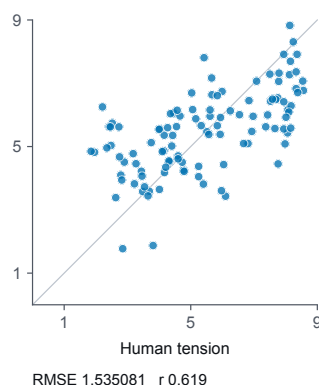


The 65-excerpt subset is nested within 356.

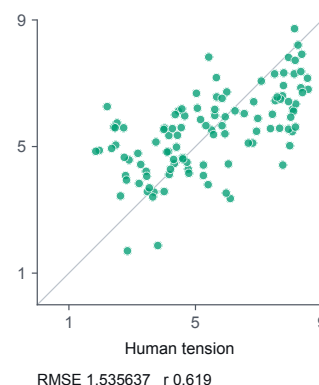
c R³ means



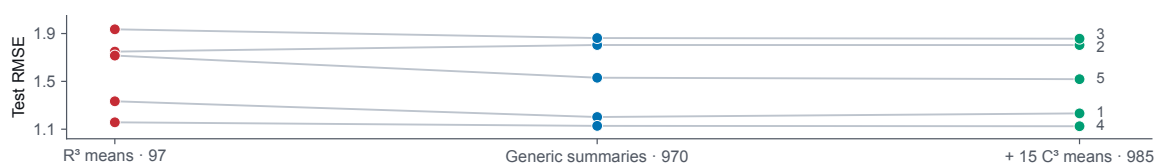
d Generic summaries



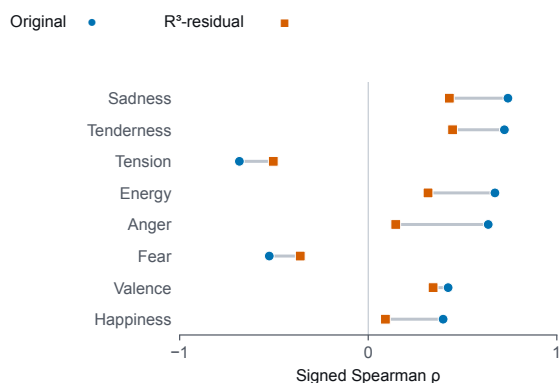
e Generic + C³ means



f Five matched test folds



g Film ratings · 110 entries



Max $|\rho|$ selected within each target pool; the pilot is not independent replication.

h Overlapping 360-entry pilot

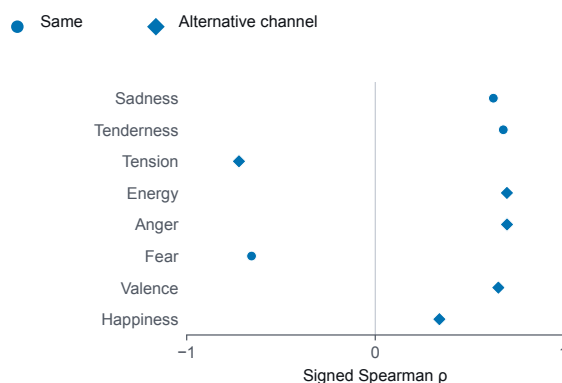


Figure 6. Excerpt-level prediction and cross-corpus arousal. (a) Five representations on 110 entries and five soundtrack-grouped folds; input counts follow each name. Acoustic 15 is the fixed comparator; generic 970 denotes acoustic/temporal descriptors, not canonical T³. RMSE is primary. (b) All four 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. This distinction becomes central in speech, where a native valence ordering and a separately fitted instruction classifier lead to different conclusions about the same recordings.

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. It does not make every arousal-labelled output interchangeable: the speech experiment uses AAC.P2 as its native scalar, whereas the historical transfer includes AAC.F1 and other channels. The response-spectrum analysis preserves the historical pool identities explicitly rather than retrospectively choosing whichever speech channel gives the most favourable association.

From correspondence to experimental contrasts

Natural-sound transformations expose a complementary property: a response can depend strongly on the reference from which it is measured. The same flute and piano endpoints reverse their displacement ordering between a birdsong and a transcription origin. Speech then shows that a multichannel representation can contain instruction information while a particular native score preserves an inverse ordering. Together, the two 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.

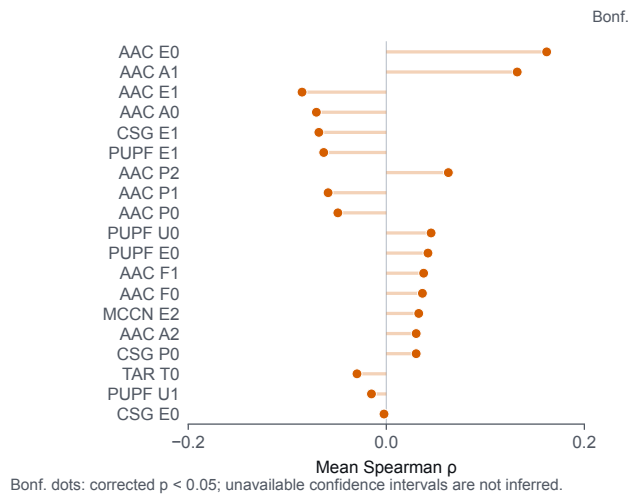
The temporal and perceptual comparisons that follow retain the same distinction. An excerpt average can correspond to a rating while losing within-phrase changes. A physiological label can organize a proposed comparison without itself supplying a physiological measurement. Figures 7 and 8 preserve both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

Compare the representation and the observation rule

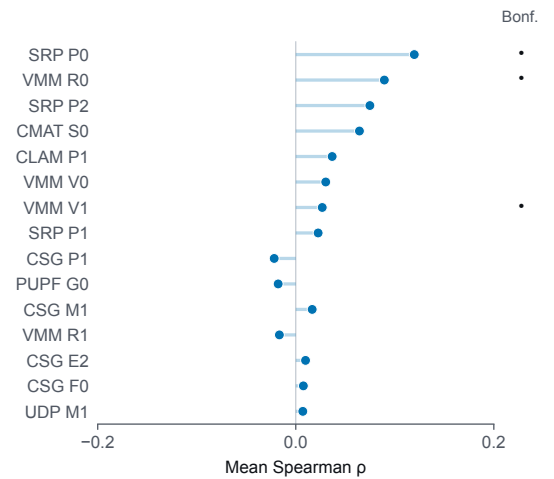
The grouped film analysis illustrates why both parts of the comparison matter. Its target is a continuous human tension rating; the speech target is one of five production instructions. The former uses nested regression and soundtrack groups, while the latter evaluates classifiers with texts held out. A numerical improvement in one cannot be carried over as evidence for the other. What transfers is the method of comparison: freeze the representation, specify the target, keep related observations together and fit the observation mapping only within training data.

Selected channels do not define the entire representation. A multivariate map can recover distinctions despite a negative native association, but it may learn a new weighting rather than verify the existing one. The speech analysis records native equations, historical selection and fitted mappings separately, preserving both fixed and nested probe results.

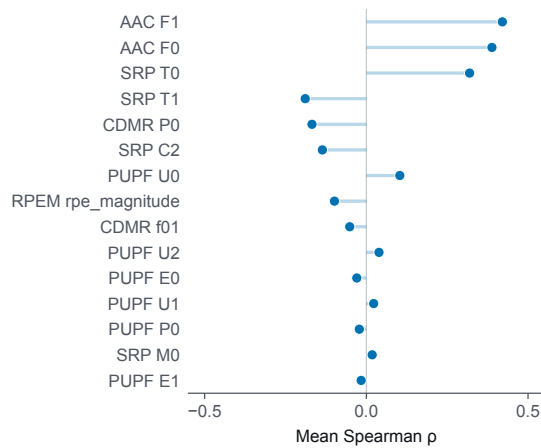
a Dynamic arousal · 19 channels



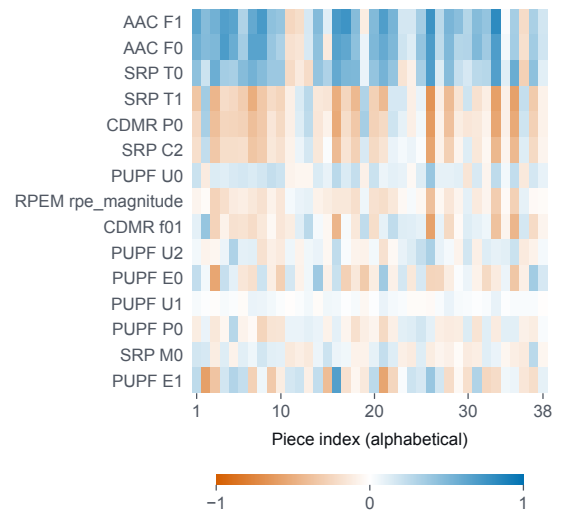
b Dynamic valence · 15 channels



c Tension · mean association



d Tension · 38 pieces



e Chill-event windows · 22 channels, three processing conditions

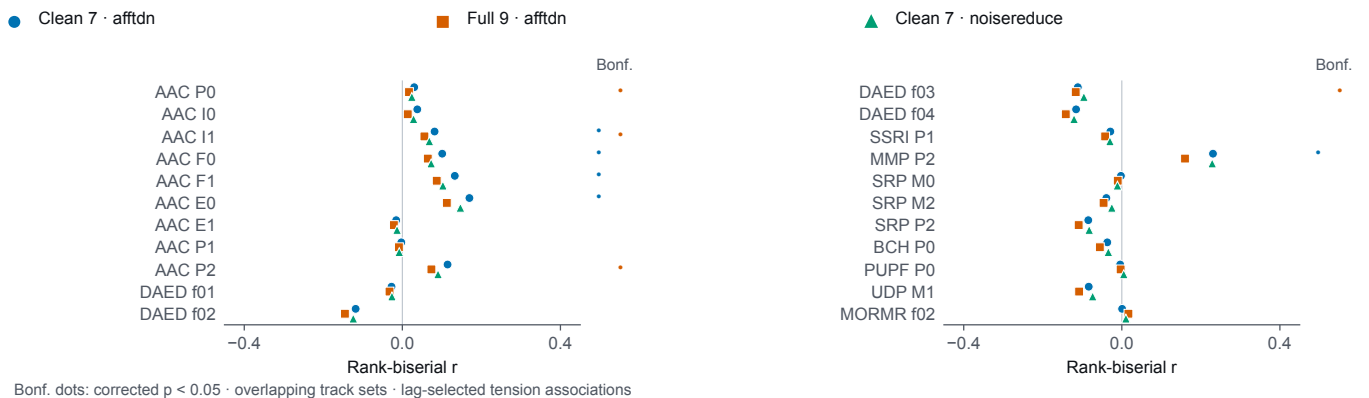


Figure 7. Time-resolved emotion, tension and reported chills. (a,b) All 19 arousal and 15 valence channels in the dynamic-emotion pilot, including negative and non-significant results. Points are saved Fisher-z means of lag-selected within-excerpt correlations. (c,d) All 15 mean tension associations 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.

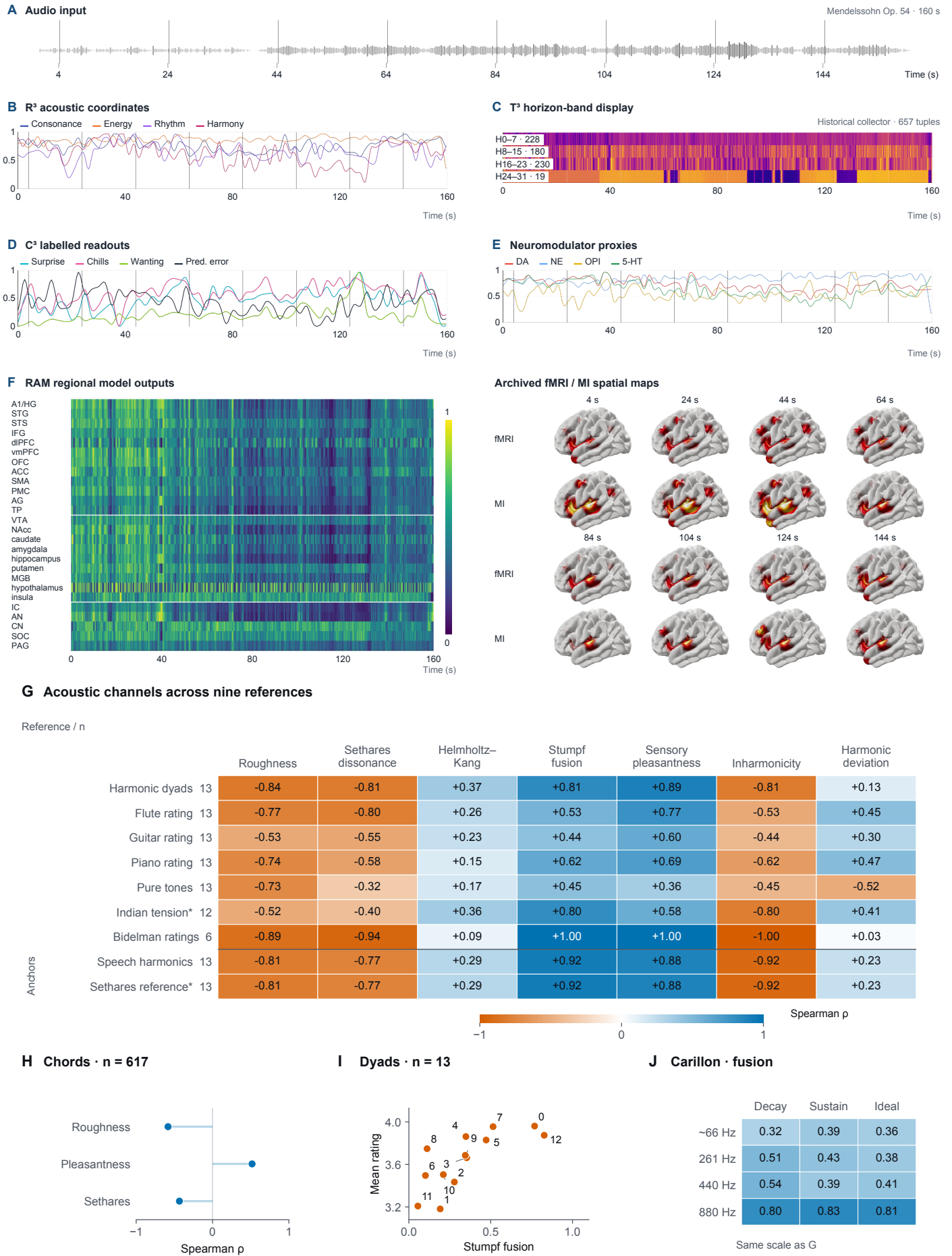


Figure 8. Acoustic representations and perceptual correspondence. (A–E) Archived 160-s Mendelssohn model displays retain source scaling; their 657-tuple collector differs from the current 644-address configuration. (F) RAM (left) and sixteen original brain views (right), pairing archived fMRI and MI 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 readout audit below provides a concrete example: a near-zero phrase mean conceals substantial signed changes within that phrase. 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 defines a route that can be inspected when another readout changes. In the speech intervention, restoring source consonance features markedly suppresses a native valence increase. That intervention concerns an executed dependency; the consonance-rating correlations concern human observations. Their conjunction identifies a plausible route for further testing without treating a successful acoustic descriptor as validation of every downstream emotion label.

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.

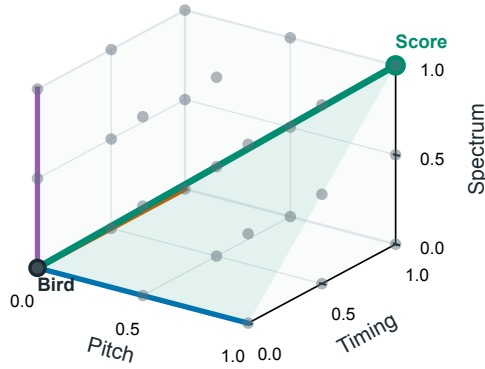
What the studies jointly establish

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 new transformation assays extend the first three beyond conventional musical excerpts and expose the consequences of source selection, clipping, feature weighting and temporal averaging. They make these choices experimentally accessible rather than leaving them implicit in a single score.

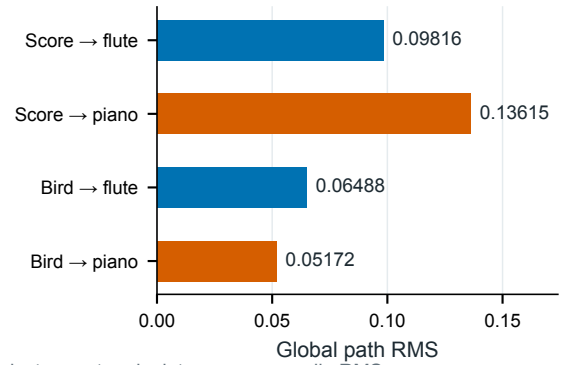
The complete comparisons also retain failures to transfer. The dynamic-emotion pilot has no Bonferroni-significant arousal channel among the 19 shown in Figure 7; an excerpt-level association therefore does not establish a temporally aligned prediction. The additional MUSIFEAST arousal family gives $\Delta R^2 = -0.016$ 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 speech plates extend this principle from existing human-data comparisons to controlled interventions on the model's inputs and readouts.

a A controlled transformation space

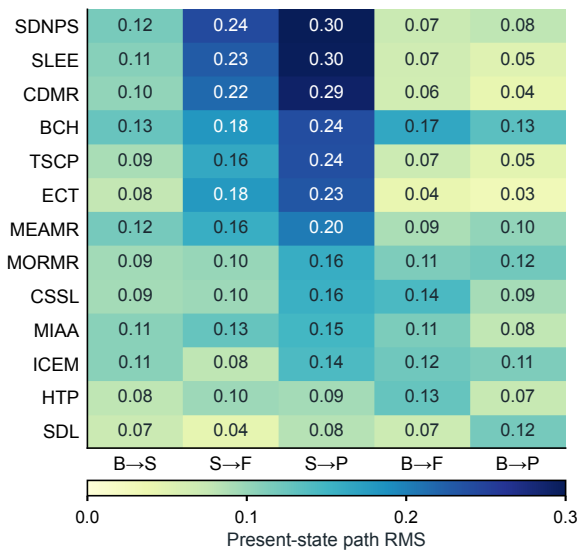


b The reference reverses the ordering

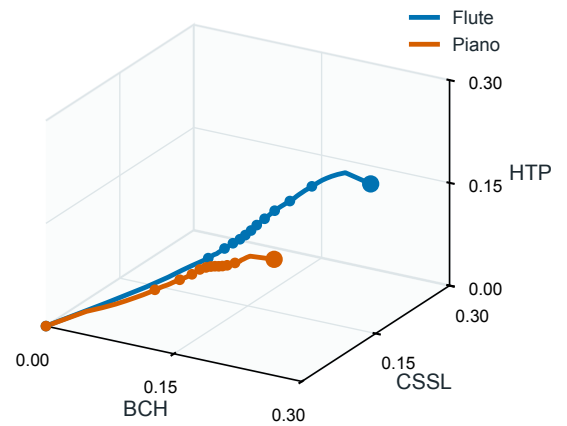


Same instrument endpoints; common audio RMS.
All 202 present-state dimensions enter the distance.

c Mechanism sensitivity depends on the path

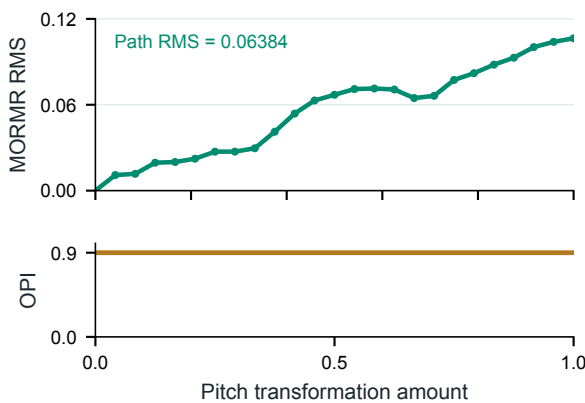


d Direct paths in named coordinates



Each axis: RMS displacement from bird.
No fitted embedding; larger markers identify $u = 1$.

e A terminal ceiling conceals change



f A mean conceals temporal change

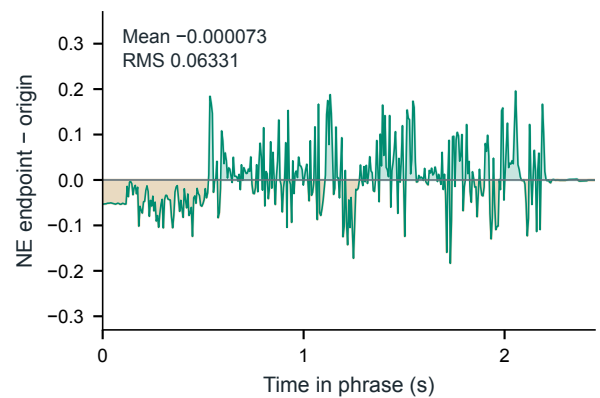


Figure 9. Reference-dependent responses and readout constraints in birdsong transformations. (a) Three renderer coordinates and their 27 factorial settings; the diagonal links bird to the operational score. (b) Exact global path RMS from score or bird origins, using identical instrument endpoints and common whole-phrase audio RMS. (c) The union of the five leading mechanisms on each of five paths (13 mechanisms); cells pool squared changes over anchors, frames and each mechanism's P dimensions. B, bird; S, score; F, flute; P, piano. (d) All 97 direct-path anchors in three named RMS coordinates; every eighth anchor is marked. These are not latent dimensions or calibrated perceptual distances. (e) On the isolated pitch path, MORMR changes while every OPI frame stays at 0.9 because its terminal source is clipped. (f) Signed NE timing-endpoint changes at identical frame indices; cancellation makes the mean small relative to framewise RMS. All renderings use fresh state. No uncertainty or significance is inferred from dependent anchors. The source study retains all 84 mechanisms, numerical-resolution controls and regional-weight analyses (Erdem, 2026a).

Natural sound: reference-dependent responses

The birdsong study processed one 2.45-s blackbird phrase through five paths linking the source, an operational 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 resulting assay measures controlled response geometry, not exposure-dependent adaptation.

With whole-phrase audio RMS matched, global present-state path displacement was 0.13615 for transcription-to-piano and 0.09816 for transcription-to-flute. From the bird, the order reversed: 0.06488 for flute and 0.05172 for piano (Figure 9b; Table 1). Endpoint RMS retained the reversal, at 0.17155 versus 0.12373 from the transcription and 0.13306 versus 0.09191 from the bird. The instrument endpoints are identical across origins. Reference and trajectory therefore determine the comparison; an endpoint has no single context-free displacement. Such a reversal can also occur in a linear representation and does not by itself demonstrate cognitive context.

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.

The leading mechanism also changed. SDNPS, SLEE and CDMR led score-origin responses; BCH led both direct bird–instrument paths. Cross-origin rank agreement fell to $\rho = 0.471$ for flute and 0.455 for piano, despite 0.950 agreement between the score-origin instruments. Spectral replacement produced greater displacement than isolated pitch or timing changes under the tested operator definitions, and factorial effects were non-additive. Refining 49 to 97 anchors stabilized ranks ($\rho = 0.99717$ – 0.99998), while midpoint errors remained 4.4–19.6% of path RMS. Ranking stability and local numerical resolution are thus separate results.

Source tracing located prominent SLEE/CDMR changes in explicit pleasantness and inverse-roughness terms and exposed discrepancies between comments and executed feature indices. The terminal audit showed another distinction: MORMR changed along the pitch path (RMS 0.06384), while OPI remained exactly 0.9 through clipping of its final source. On the timing path, NE’s signed endpoint mean was -0.000073 , but framewise RMS was 0.06331. Regional concordance decreased from $r = 0.773$ to 0.638 after division by incoming absolute weight sums. Readout rules can therefore conceal or rescale changes present in the underlying representation.

Speech: information beyond a native affect score

The speech study tests transfer of this music-oriented architecture to vocal expression (Erdem, 2026b). Shared acoustic-affect cues motivate that comparison, while different spectral–temporal dependencies caution against assuming identical mappings (Juslin and Laukka, 2003; Albouy et al., 2020). The corpus contains 342 recordings from 172 generation requests. Its primary factorial crosses three texts, two delivery modes, five instructions and

three requests. Alternatives remain grouped; production instructions, not human emotion ratings, supply the labels.

On three held-out-text folds, fixed-ridge speech accuracy was 35.6% for a fitted native affect pair, 60.0% for the 82-channel historical pool union and 68.9% for all 889 C³ dimensions; 15 acoustic descriptors reached 57.8% (Figure 10b). Nested regularization changed the ordering: union and C³ reached 46.7% and 48.9%. Acoustic-plus-pool prediction reached 64.4% with fixed and 62.2% with nested tuning, against 57.8% for acoustics. The observed complement is promising, but three text folds do not establish a general predictive advantage. A fitted probe tests accessible information, not the correctness of the native interpretation (Hewitt and Liang, 2019).

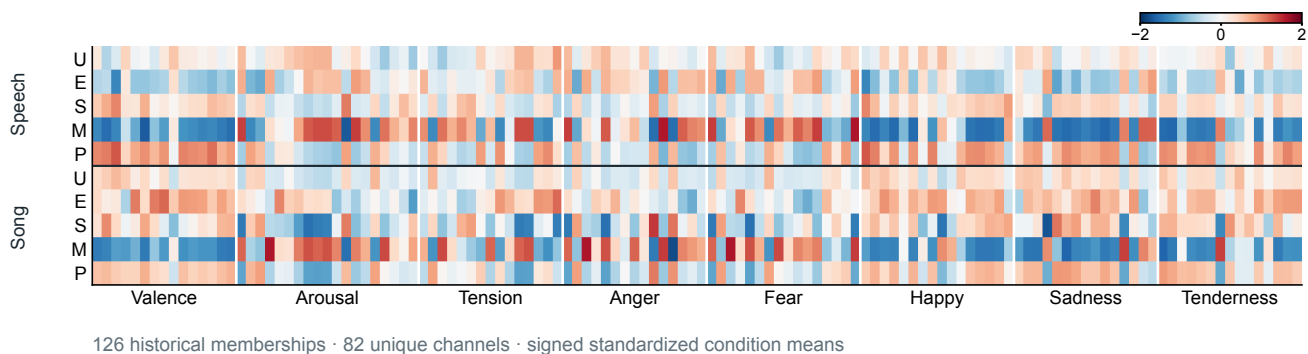
The response spectrum also separates geometry from direction. Speech–song distance-rank agreement was 0.855 in the tension pool, 0.733 in arousal and 0.370 in valence. Native valence nevertheless assigned excited speech lower values than sad speech in every text. In six selected sources, instrument-derived upper-harmonic transfer and residual attenuation increased mean valence from 0.139 to 0.435–0.437, yet all 588 primary excited-minus-sad anchor contrasts stayed negative. Restoring the source consonance block suppressed 94.2–94.8% of the mean absolute increase. This locates a model dependency without repairing the native ordering.

The controls further separate these effects. Envelope transfer alone increased native valence by about 0.015, residual removal alone by 0.221, and their joint application by 0.297–0.299. Their positive interaction was 0.061–0.062 model units. Arousal followed a different dependence: excited exceeded sad in all 294 primary native-level comparisons, but only 94 of 294 common-RMS comparisons, with negative common-RMS endpoint contrasts. Level matching therefore changes the interpretation of an apparently consistent scalar direction.

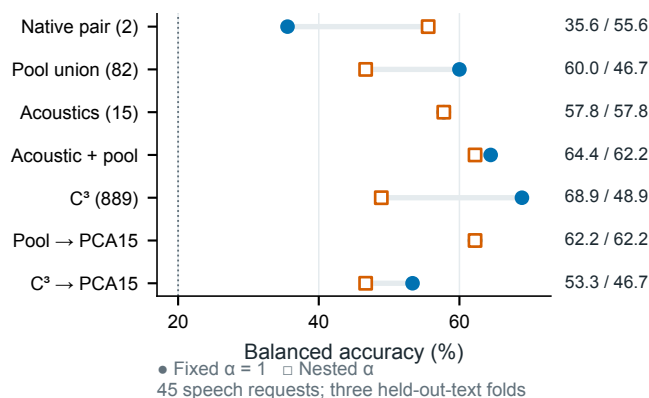
The dimension-matched comparison is similarly informative. Pool PCA15 reached 62.2% speech accuracy under both fitting procedures, whereas C³ PCA15 reached 53.3% with fixed and 46.7% with nested regularization. More channels do not guarantee more transferable information. At fixed regularization, adding the historical pool to acoustics improved five held-out requests and worsened two; nested tuning improved six and worsened four. These paired outcomes retain the actual scale of the observed complement. They identify a concrete readout to test on new voices and texts, while the existing three-text analysis remains conditional on this corpus. All alternative procedures and negative comparisons are retained in the speech companion.

Multichannel contrasts changed as well: mean source-to-endpoint cosine was about 0.22 for arousal and 0.64 for valence at native level. Near-endpoint waveform differences of -86.7 to -81.8 dB RMS accompanied valence gaps up to 0.1073; one of 96 additional boundary contrasts was weakly positive. The primary result is therefore distributed instruction information plus a traceable, numerically sensitive readout. It establishes neither emotion recognition in people nor selective removal of consonants. Together with birdsong, it shows why reference, representation, readout and stability must be examined jointly.

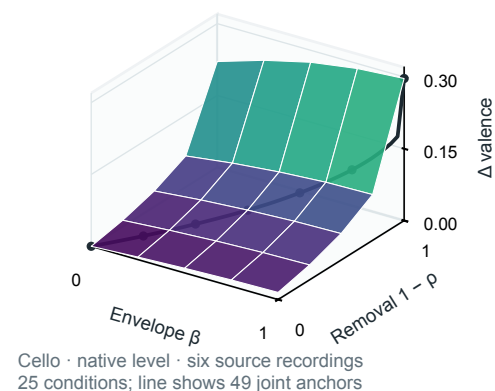
a Affect is a response pattern, not one score



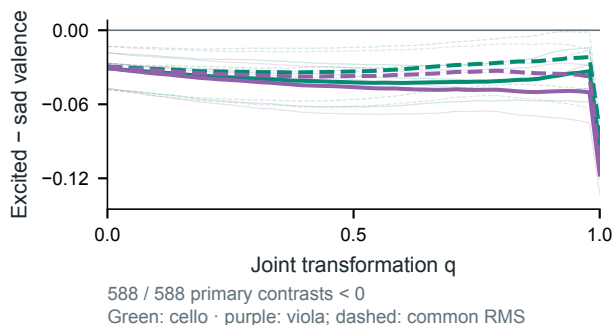
b Fitted instruction information



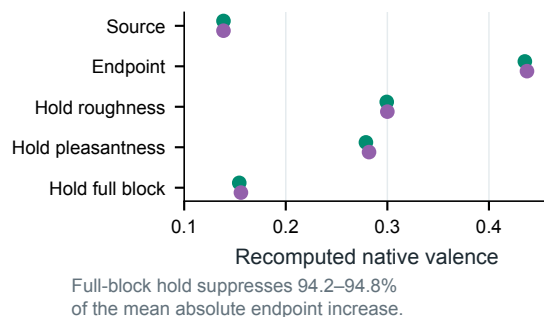
c Two controlled acoustic changes



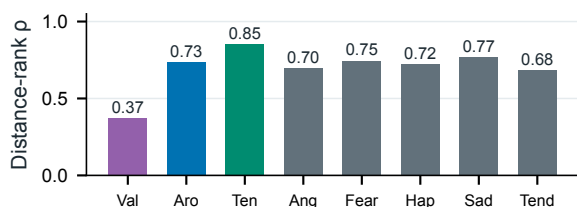
d Native ordering is not repaired



e Locate the source of the score shift



f Shared geometry varies across pools



g Multichannel contrasts rotate

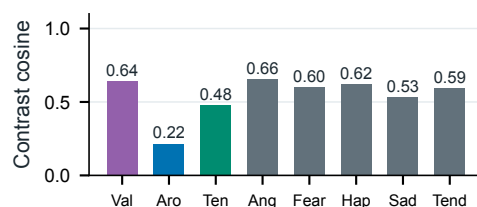


Figure 10. Speech distinguishes distributed information, native score direction and internal dependence. (a) Historical pool profiles: five instructions in speech and song, standardized over 90 original main requests. U/E/S/M/P denote untagged/excited/sad/mad/peaceful; colour saturates at ± 2 . Overlapping memberships are not independent channels. (b) Balanced instruction accuracy for identical held-out-text folds; printed pairs give fixed/nested ridge results. PCA15 is fitted inside training folds. (c) Cello/native-level valence change across the 25-condition envelope/residual grid; the line traces 49 evaluated joint anchors. Facets and lines are visual guides. (d) Excited-minus-sad native valence: thin curves show texts, thick curves their mean; colours show envelopes and dashes common RMS. All 588 primary contrasts are negative; a separate boundary sweep has one weakly positive contrast. (e) Endpoint valence after restoring source features and recomputing T³/C³. Suppression of the absolute increase does not repair the ordering. (f) Speech–song rank agreement of ten dependent condition distances per pool. (g) Source-to-endpoint excited–sad contrast cosine, averaged over three texts and two envelopes at native level. Prompt labels are production instructions; these panels do not measure human emotion or calibrated affect probabilities (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	DGTP F3 U	Domain-General Temporal Processing Rathcke 2024 · Grah 2007	DAED F6 H	Dopamine Anticipation-Experience Dissociation Salimpoor 2011 · Gold 2023
	CSG F1 A	Consonance-Saliency Gradient Bravo 2017 · Sarasso 2019	SDL F3 U	Saliency-Dependent Lateralization Haiduk 2024 · Zatorre 2022	MORMR F6 H	mu-Opioid Receptor Music Reward Putkinen 2025 · Salimpoor 2011
	MIAA F1 U	Musical Imagery Auditory Activation Kraemer 2005 · Halpern 2004	NEWMD F3 U	Neural Entrainment-Working Memory Dissociation Noboa 2025	RPEM F6 P	Reward Prediction Error in Music Gold 2019a · Gold 2023
	MPG F1 U	Melodic Processing Gradient Rupp 2022 · Patterson 2002	PWSM F3 U	Precision-Weighted Saliency Model Basinski 2025 · Millidge 2022	IUCP F6 U	Inverted-U Complexity Preference Gold 2019b · Gold 2023
	PNH F1 U	Pythagorean Neural Hierarchy Bidelman 2009 · Tabas 2019	MEAMN F4 U	Music-Evoked Autobiographical Memory Network Janata 2009 · Sakakibara 2025	MCCN F6 U	Musical Chills Cortical Network Chabin 2020 · Putkinen 2025
	SDNPS F1 U	Stimulus-Dependent Neural Pitch Saliency Cousineau 2015 · Penagos 2004	MMP F4 CH	Musical Mnemonic Preservation Jacobsen 2015 · Fang 2017	MEAMR F6 U	Music-Evoked Autobiographical Memory Reward Janata 2009 · Salimpoor 2011
	SDED F1 U	Sensory Dissonance Early Detection Crespo-Bojorque 2018 · Fishman 2001	RASN F4 U	Rhythmic Auditory Stimulation Neuroplasticity Thaut 2015 · Grah 2007	SSRI F6 U	Social Synchrony Reward Integration Kokal 2011 · Bigand 2025
	TPIO F1 U	Timbre Perception-Imagery Overlap Halpern 2004 · Zatorre 2005	PMIM F4 U	Predictive Memory Integration Model Koelsch 2000 · Koelsch 2009	LDAC F6 U	Liking-Dependent Auditory Cortex Gold 2023 · Cheung 2019
	TPRD F1 U	Tonotopy-Pitch Representation Dissociation Allen 2022 · Norman-Haignere 2013	OII F4 U	Oscillatory Intelligence Integration Bruzzone 2022 · Samies 2022	IOTMS F6 U	Individual Opioid Tone Music Sensitivity Putkinen 2025 · Salimpoor 2011
	PSCL F1 U	Pitch Saliency in Cortical Lateralization Penagos 2004 · Allen 2022	HCMC F4 U	Hippocampal-Cortical Memory Circuit Squire 1995 · McClelland 1995	SSPS F6 U	Saddle-Shaped Preference Surface Cheung 2019 · Gold 2019b
F2 Prediction	PCCR F1 U	Pitch Chroma Cortical Representation Briley 2013 · Allen 2022	RIRI F4 U	RAS-Intelligent Rehabilitation Integration Thaut 2015 · Harrison 2025	PEOM F7 U	Period Entrainment Optimization Model Thaut 2015 · Thaut 1998
	HTP F2 H	Hierarchical Temporal Prediction Norman-Haignere 2022 · Bonetti 2024	MSPBA F4 U	Musical Syntax Processing in Broca's Area Maess 2001 · Kim 2021	MSR F7 U	Musician Sensorimotor Reorganization Bangert 2006 · Lahav 2007
	SPH F2 U	Spatiotemporal Prediction Hierarchy Bonetti 2024 · Fernandez-Rubio 2022	VRIAP F4 U	VR-Integrated Analgesia Paradigm Deng 2025	GSSM F7 U	Gait-Synchronized Stimulation Model Yamashita 2025
	ICEM F2 CH	Information Content Emotion Model Egermann 2013 · Cheung 2019	CMAPEC F4 U	Cross-Modal Action-Perception Common Code Lahav 2007 · Bangert 2006	HMCE F7 U	Hierarchical Musical Context Encoding Mischler 2025 · Norman-Haignere 2022
	PWUP F2 C	Precision-Weighted Uncertainty Processing Mencke 2019 · Cheung 2019	DMMS F4 U	Developmental Music Memory Scaffold Partanen 2022 · Qiu 2025	HGSIC F7 U	Hierarchical Groove State Integration Circuit Potes 2012 · Grah 2007
	WMED F2 U	Working Memory-Entrainment Dissociation Noboa 2025 · Yuan 2025	CSSL F4 U	Cross-Species Song Learning Burchardt 2025	ASAP F7 U	Action Simulation for Auditory Prediction Ross 2022 · Patel 2014
	UDP F2 U	Uncertainty-Driven Pleasure Mencke 2019 · Gold 2019b	CDEM F4 U	Context-Dependent Emotional Memory Sachs 2025	DDSMI F7 U	Dyadic Dance Social Motor Integration Bigand 2025 · Kohler 2025
	CHPI F2 C	Cross-Modal Harmonic Predictive Integration Moller 2021 · Cheung 2019	SRP F5 PA	Striatal Reward Prediction Salimpoor 2011 · Salimpoor 2013	VRMSME F7 U	VR Music Stimulation Motor Enhancement Liang 2025
	IGFE F2 U	Individual Gamma Frequency Enhancement Yokota 2025 · Bolland 2025	AAC F5 CPA	Autonomic-Arousal Circuit Egermann 2013 · Salimpoor 2011	SPMC F7 U	SMA-Premotor-M1 Motor Circuit Grah 2007 · Hoddinott 2024
	MAA F2 U	Multifactorial Attention Appreciation Mencke 2019 · Gold 2019b	VMM F5 A	Valence-Mode Mapping Mitterschiffthaler 2007 · Koelsch 2006	NSCP F7 U	Neural Synchrony Commercial Prediction Leeuwis 2021
F3 Attention	PSH F2 U	Prediction Silencing Hypothesis Millidge 2022 · Aukstulewicz 2016	PUPF F5 PA	Pleasure-Uncertainty Prediction Function Cheung 2019 · Egermann 2013	CTBB F7 U	Cerebellar Theta-Burst Balance Sansare 2025 · Okada 2022
	SNEM F3 U	Sensory Novelty and Expectation Model Nozaradan 2011 · Nozaradan 2012	CLAM F5 U	Closed-Loop Affective Modulation Ehrlich 2019 · Daly 2016	STC F7 U	Singing Training Connectivity Zamorano 2023 · Kieber 2013
	IACM F3 U	Inharmonicity-Attention Capture Model Basinski 2025 · Koelsch 1999	MAD F5 U	Musical Anhedonia Disconnection Loui 2017 · Martinez-Molina 2016	EDNR F8 U	Expertise-Dependent Network Reorganization Paraskevopoulos 2022 · Leopold 2021
	BARM F3 U	Brainstem Auditory Response Modulation Rathcke 2024 · Grah 2007	NEMAC F5 U	Nostalgia-Evoked Memory-Affect Circuit Sakakibara 2025 · Barrett 2010	TSCTP F8 U	Temporal Synaptic Consolidation Processor Pantev 2001 · Bellmann 2024
	STANM F3 U	Spectrotemporal Attention Network Model Haiduk 2024 · Zatorre 2022	STAI F5 U	Spectral-Temporal Aesthetic Integration Kim 2019 · Alluri 2012	CDMR F8 P	Context-Dependent Mismatch Response Rupp 2022 · Crespo-Bojorque 2018
	AACM F3 U	Aesthetic-Attention Coupling Model Sarasso 2019 · Sarasso 2021	DAP F5 C	Developmental Affective Plasticity Scholkman 2024 · Trainor 2012	SLEE F8 U	Synaptic Long-term Encoding Engine Paraskevopoulos 2022 · Porfiri 2025
	AMSS F3 U	Attention-Modulated Stream Segregation Hausfeld 2021 · Wikman 2025	CMAT F5 A	Cross-Modal Affective Transfer Spence 2011	ESME F8 U	Error-Signal Modulated Encoding Koelsch 1999 · Vuust 2012
	ETAM F3 U	Entrainment, Tempo and Attention Modulation Hausfeld 2021 · Doelling 2015	TAR F5 U	Therapeutic Affective Resonance Kheirkhah 2025 · Ehrlich 2019	ECT F8 U	Expertise Compartmentalization Trade-off Paraskevopoulos 2022 · Porfiri 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 $p = 0.691$; A1 $p = 0.533$ (356 excerpts).	Synchronized physiology and acoustic controls for an autonomic claim.
	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 $p = 0.676$ (356 excerpts). This is not a pleasure test.	Pleasure-specific targets and alternatives for the uncertainty interpretation.
F1 Sensory	CSG Consonance-Saliency Gradient	F1 Bravo et al. (2017) Sarasso et al. (2019) Interval acoustics relate to affective appraisal and attention.	Consonance cues supply saliency- and affect-labelled model outputs.	Arousal: P0 $p = 0.531$ (356 excerpts). Saliency itself is not measured.	Attention/saliency 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 and two transformation studies demonstrate this use across temporal, spectral and readout interventions. 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 predictive evidence for that task. The speech results add a complementary case: distributed instruction information is accessible, but its measured advantage depends on representation and fitting choices. Neither result establishes that a larger repertoire is always preferable. Efficiency, upstream cost and the time required to develop or revise a model remain to be measured.

The transformation studies show why the readout operator belongs inside the scientific account. Birdsong displacement reverses across reference sounds, a clipped terminal channel can remain constant while a mechanism changes, and averaging can cancel substantial temporal differences. In speech, instruction information is distributed across channels even when a native valence contrast has the inverse sign. Holding a source feature block suppresses most of the absolute score rise, exposing a computational route that a scalar correlation would not identify. These are demonstrated properties of the implemented models. They motivate specific revisions and comparisons without requiring a claim that the labels already describe human mechanisms.

Reusable components and explicit graphs have precedents. Es-sentia 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 while refining its interpretation. Their common finding is that response magnitude alone is insufficient: the reference and transformation determine a contrast, and the observation rule determines what survives into a score. Birdsong establishes this for controlled sound trajectories; speech demonstrates it for intended expression and an explicit feature intervention. The assays share a logic rather than a pooled effect size. Birdsong RMS uses framewise P-layer tensors, whereas speech probes and morph profiles use recording summaries and different channel selections.

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). This makes domain-specific cue weighting a substantive hypothesis for MI's native-score behaviour. It is not the only explanation: miscalibration, numerical sensitivity and a mismatch among intended, perceived and felt affect remain alternatives. Residual attenuation does not selectively remove consonants, and worded song retains phonetic structure. Segment-aligned controls, new voices and independent perceptual targets would distinguish these possibilities.

The positive result is consequently precise: the representation retains instruction-sensitive structure, and controlled interventions identify a substantial route into its native valence calculation. The inverse primary ordering persists, while the full response changes. Neither fact cancels the other. This separation of information from interpretation follows the logic of diagnostic probing (Hewitt and Liang, 2019); the numerical boundary findings add a prerequisite for interpreting smooth perceptual transitions. More generally, model-metamer work shows why task performance alone does not establish human-like invariances (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 the birdsong assay, matched R^3 , log-mel and envelope distances would locate where reference

ordering arises; component interventions and independent source phrases would test attribution and generality. 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 T³ 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](#), 15 October 2026 editorial edition, and [speech](#), 15 October 2026 response-spectrum edition. The exact public main and supplementary PDFs, input hashes and figure-source arrays accompany this integrated revision. Their source archives retain all mechanisms, signed outputs, negative comparisons and diagnostic controls. The earlier Zenodo DOI does not

identify these experiments. The original R3 Scientific Supplement remains intact as the record for the historical architecture and human-data analyses; the companions document the later readout and speech work. Figure numbers in the original supplement retain the R3 sequence; its Figure 9 is Figure 11 here. No new participants, audio generation or engine inference were added for this edition.

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-Mirallès, A., Quiroga-Martínez, 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. Instruction information and spectral sensitivity in a musical model of speech and song. *Research manuscript*, 15 October 2026 response-spectrum edition (2026b). [Source](#)
- Erdem, A. Reference-dependent responses to birdsong-to-music transformations. *Research manuscript*, 15 October 2026 editorial edition (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. & Przyssinda, 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](#)
- Marjeh, 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., Vuhan, 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](#)
- Scholkman, 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., Nääätänen, 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 R³ 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 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 ECT[*f*04], OPI= 0.90 IOTMS[*E*0] and 5HT= 0.30 BCH[*P*0]. 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 15 October birdsong companion and its supplement (Erdem, 2026a); the original architectural S8 remains a record of the earlier assay.

Speech design and instruction comparisons

Speech and song were generated with one displayed voice and fixed intended words. The main factorial comprises 90 requests and 179 available alternatives; overlapping extension panels bring the archive to 172 requests and 342 recordings. The interface displayed Eleven v4 and Anna—Young Female Conversational Voice, with stability 0.5 and similarity 0.75. Prompts, production order and file hashes define the archived stimuli; an exact check-point and seed were unavailable. Six transcript disagreements were retained. Available alternatives are averaged within request, and texts receive equal weight. Prompt labels do not provide continuous affect ratings. Matched-statement human speech/song corpora such as RAVDESS provide an independent target for future comparisons (Livingstone and Russo, 2018), not additional observations in this experiment.

Both speech studies retain the same frozen adapter and 855-file engine pin, with fresh state for every full recording. The adapter uses 44.1-kHz mono float32 input, a 2,048-point FFT, hop 256, 128 mel bands, $\log(1+\text{power})$ and recording-wide maximum normalization. The waveform also reaches acoustic extraction. Native valence is VMM.C1 and native arousal is AAC.P2; these are specified outputs, not the full affect representation. Full-file summaries retain 889 C³ dimensions, 131 unstretched combinations, 26 regional projections and four modulator proxies. The bird-song P-code slice contains 202 dimensions, whereas the speech present-name slice contains 195; these are different registry selections, despite matching ordered 889-channel names. Their RMS quantities are not pooled across experiments.

Historical pools, geometry and fitted speech probes

The post hoc response-spectrum amendment transfers eight historical pool definitions by exact identifier, matching 126 of 128 memberships and 82 of 84 unique channels. Two unavailable PNH dimensions remain missing. Identifier agreement does not establish identical implementations across engine versions. Descriptive condition profiles standardize over the 90 original main requests. Geometry compares the ten dependent pairwise distances among five instruction centroids between speech and song, following the representational-comparison logic of Kriegeskorte et al. (2008). These rank correlations do not use listener ratings. Contrast orientation is a cosine between excited-minus-sad vectors, computed in raw within-pool model units.

Five-class probes retain three held-out-text folds, 45 requests per delivery and nominal 20% chance. Imputation, standardization, near-constant filtering and feature-count scaling are fitted within training data. The fixed ridge uses $\alpha = 1$; nested tuning selects among 0.1, 1, 10, 100, 1,000 and 10,000 on the two remaining texts, with ties favouring larger values. PCA15 is fitted within each relevant training fold, without whitening. The 15-descriptor acoustic baseline covers level, duration, pitch, spectral and envelope properties. These classification probes adapt the grouped-readout logic used for film tension; they do not replicate its continuous human target or initial-frame omission. Small text counts constrain precision (Varoquaux, 2018). Fixed, nested, dimension-matched and paired-error results are all retained in the companion. No new superiority test is assigned to the added comparison.

The original speech analysis retains 556 contrasts and 2,919,000 scalar effects. A common family of 318,150 eligible tests yields

9,689 BH discoveries and none under global BY correction. Separate restricted label-permutation tests support instruction information for selected decoders, including fixed C³ speech, but do not test C³ against acoustics. The 296 response-spectrum summaries are descriptive additions and leave those original correction families unchanged. This preserves the distinction between a label-null test and an incremental-prediction test.

Speech spectral and feature interventions

Six sources are the first take and first available alternative for excited and sad instructions in each text, chosen without transformed MI outcomes. Their 4.99–6.43-s waveforms retain source phase, timing and sample count. An upper-harmonic envelope is interpolated towards a cello- or viola-derived template while the residual outside a pitch-dependent harmonic mask is attenuated independently: $y(\beta, \rho) = g[H_\beta + \rho R]$. Lower harmonics are protected by a raised-cosine transition over orders 2.5–3.5. There is no explicit octave shift, time warping or note/rhythm quantization. The independent pitch audit and its remaining deviations are retained. The residual is defined by a signal mask, not a phonetic segmentation.

The joint path sets $\beta = q$ and $\rho = 1 - q$ at 49 anchors. A 5×5 factorial grid, two envelopes and native/common-RMS levels yield 1,596 deduplicated primary evaluations. The common-RMS target is -37.386992 dBFS, distinct from the original corpus's level control. Every complete recording is evaluated with fresh state. Feature holds replace endpoint roughness, pleasantness or the seven-feature consonance block with original-source trajectories and recompute T³ and C³. Seventy-two interventions and 24 no-op checks locate an internal dependency; hybrid feature vectors need not correspond to realizable sounds. Suppression percentages divide the reduction in mean absolute endpoint rise by the original mean rise; they are not explained-variance or human mediation estimates.

Eight added boundary levels yield 192 evaluations and 96 paired contrasts; one contrast is weakly positive at $q = 0.995$, then negative at the endpoint. Separate renderer and peak-interpolation controls retain their own provenance. The complete analytic archive has 1,914 full-file evaluations and 1,836 unique decoded waveforms. Anchors are dependent transforms of six sources, so no anchor-level confidence intervals or population tests are introduced. Figure 10 replots archived results; connecting curves and surface facets are visual guides. No new engine execution or statistical fitting was performed for this integrated edition.

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 R³ means; 970 acoustic/temporal descriptors; those 970 descriptors plus fifteen C³ means; the fifteen C³ means alone; two training-fold PCA15 representations; a fixed fifteen-descriptor acoustic subset; and a training-target-mean reference. Each R³ 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 C³ 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.