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 presents an executable architecture for developing computational hypotheses about music listening. Its primary contribution is to organise acoustic coordinates, requested temporal operations and listening-related calculations so that their assumptions and dependencies can be inspected and revised. In a recorded execution, a targeted temporal intervention altered only outputs within the declared route, while reconstruction recovered a selected output from saved inputs. Exploratory readouts predicted film tension, with errors close to a fixed acoustic subset; selected arousal associations transferred across corpora. Together, these demonstrations provide an executable proof of concept for inspecting and revising selected computations within the architecture. Cognitive and physiological interpretations remain hypotheses, and independent reconstruction requires implementation records available through controlled access.

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

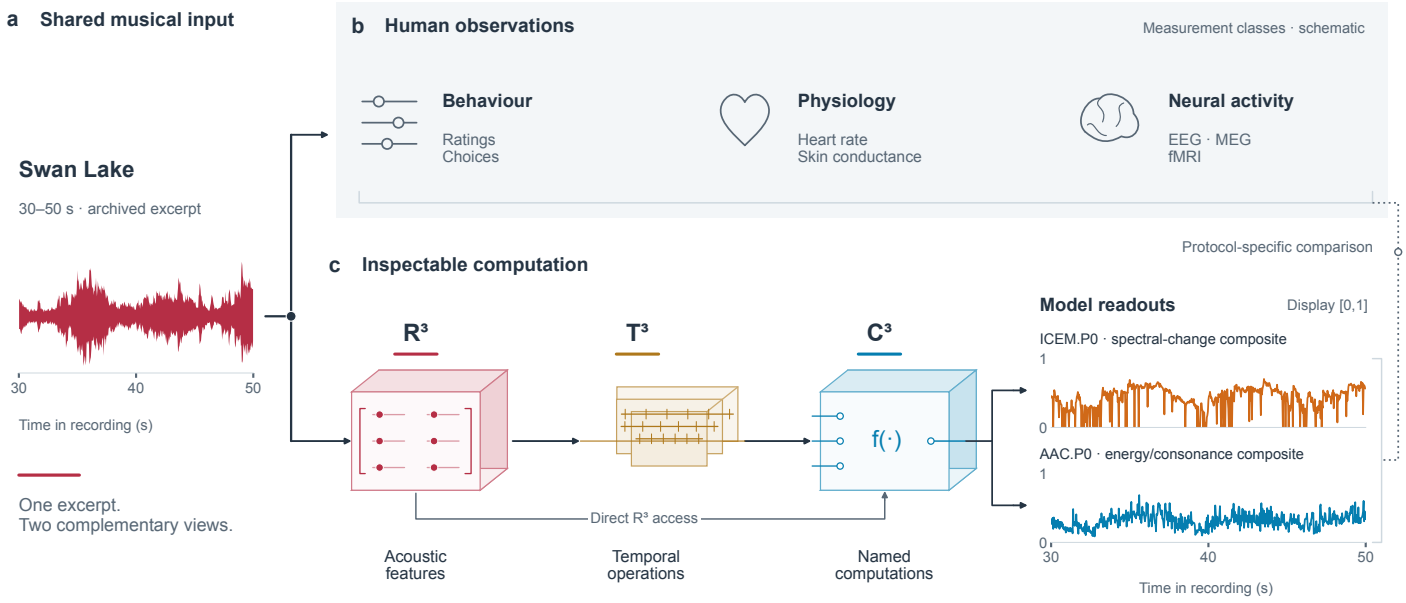


Figure 1. Music as a setting for inspectable computation. The Swan Lake excerpt supplies both paths. Observation symbols denote measurement classes, not collected participant data. R³ features enter T³ operations and C³ calculations, including direct R³ access. Depth is schematic. Curves retain their 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 aspects of sound enter a calculation, how much context it uses and how it depends on other computations. Musical Intelligence addresses this model-building problem.

Formal models make theoretical assumptions explicit and expose their consequences (Guest and Martin, 2021). Cognitive architectures extend this approach to the organisation of components and their interactions (Langley et al., 2009; Laird et al., 2017). MI applies this architectural perspective to computational hypotheses about music listening.

Audio libraries provide reusable sound-analysis algorithms (Bogdanov et al., 2013); probabilistic models specify musical expectation (Pearce and Wiggins, 2012); and TenseMusic predicts continuous tension from audio (Barchet et al., 2024). MI organises acoustic features, temporal context and listening-related calculations within a shared execution. Its primary contribution is an architecture in which these choices can be inspected together and individually revised.

R³ describes sound as named acoustic coordinates. T³ specifies temporal operations requested on those coordinates. C³ combines acoustic and temporal inputs into calculations expressing listening hypotheses. A fixed configuration gives deterministic outputs; saved inputs support reconstruction, and an altered request tests downstream consequences. Cognitive, regional and neuromodulator labels identify proposed interpretations to be assessed against observations.

Architectural evaluation reconstructs an output and follows a controlled change through its consumers. Empirical case studies then compare selected readouts with human observations. Fitted ridge readouts predict film tension with errors close to a fixed acoustic subset, while direct arousal ranks transfer across corpora. The engine stays fixed within each comparison; fitting estimates the observation mapping. Together, these demonstrations assess the architecture's use for specifying hypotheses, examining computations and relating selected outputs to observations.

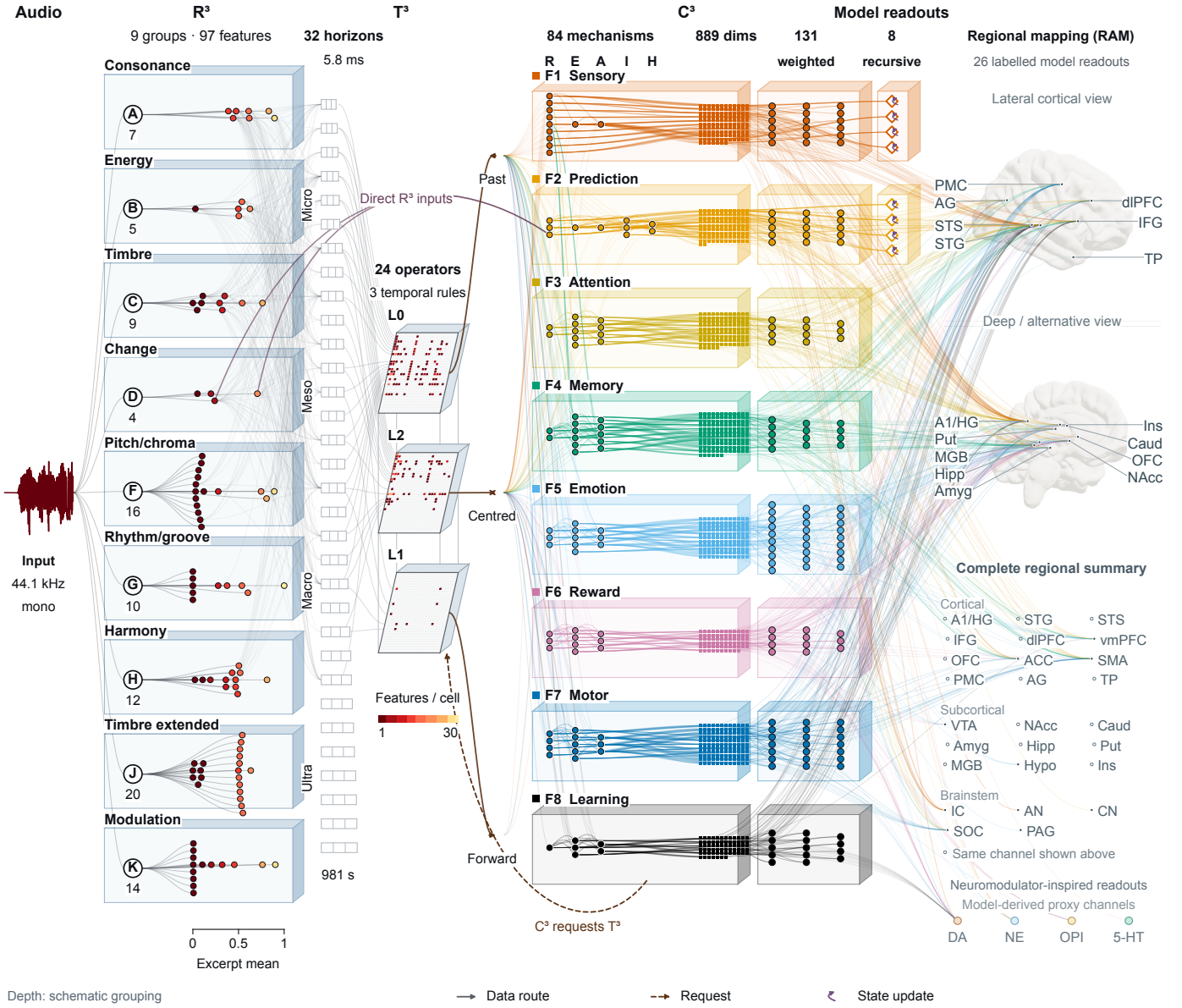


Figure 2. Musical Intelligence architecture. The current configuration contains 84 mechanisms, 889 output dimensions, 131 weighted readouts and eight selected recursive classes. Surfaces group computations; depth is schematic. Solid paths show data routes, dashed arrows requests and loops local state.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

2. Architecture

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

Configuration and observation

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

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

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

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

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

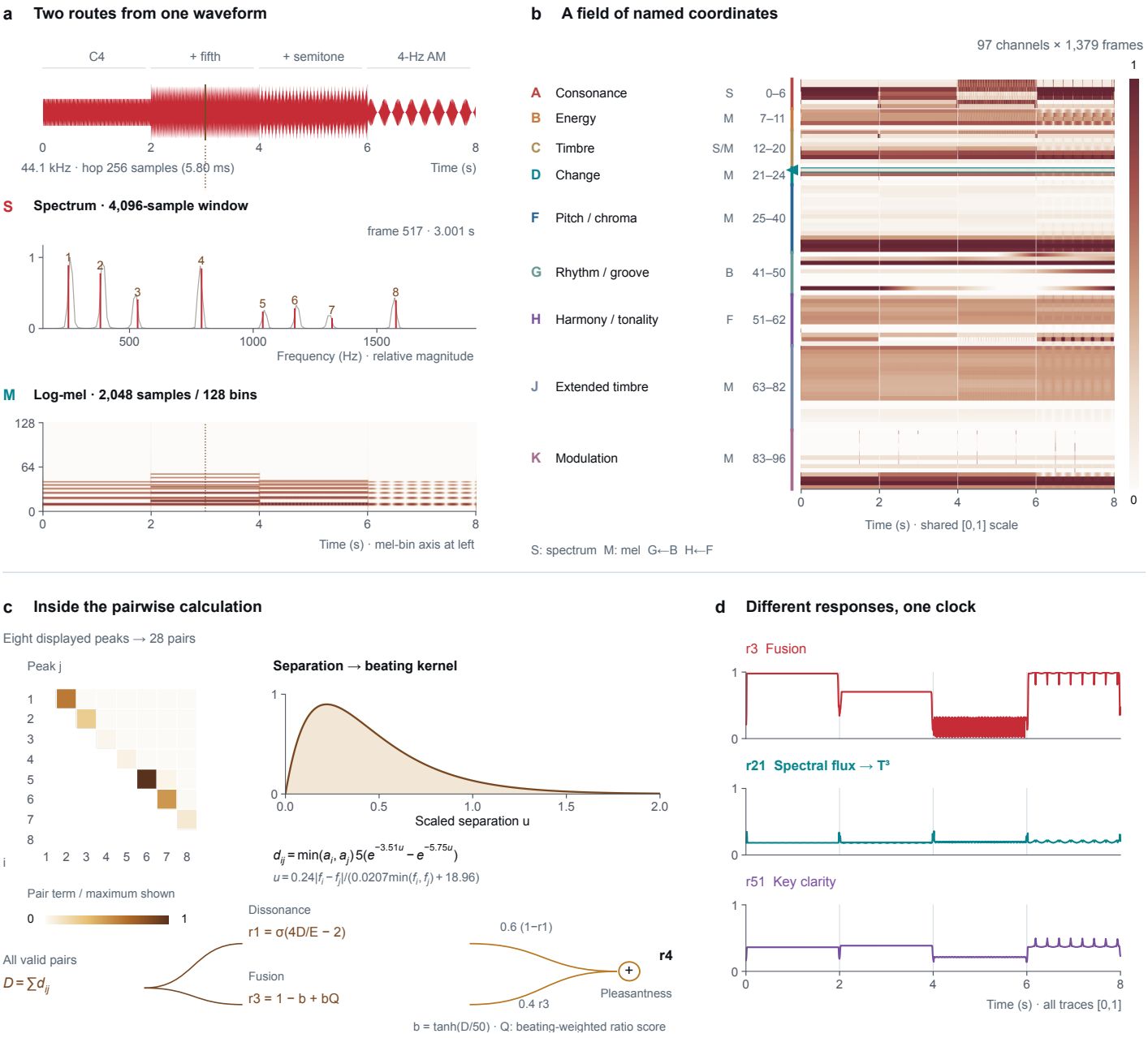


Figure 3. From sound to named acoustic coordinates. (a) The preserved 8-s waveform, saved spectrum at 3.001 s and log-mel field distinguish spectral (S) and mel (M) routes; window lengths differ from the 256-sample hop. (b) All 97 channels and 1,379 frames retain their [0,1] scale. S supplies A and C's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

Acoustic coordinates

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

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

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

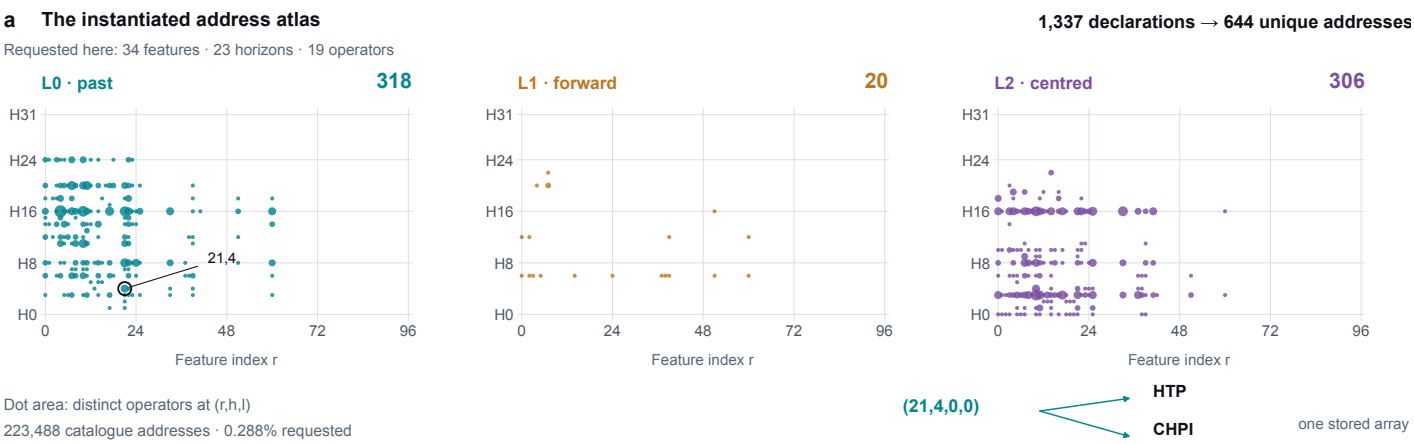


Figure 4. A temporal address opens into a support geometry and a specific operation. (a) The canonical 1,337 declarations reduce to 644 unique addresses. Maps preserve feature--horizon positions; dot area counts distinct operators at each position. Law totals count addresses. HTP and CHPI share $(21,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

Requested temporal operations

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

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

C³ · ICEM computation and readouts

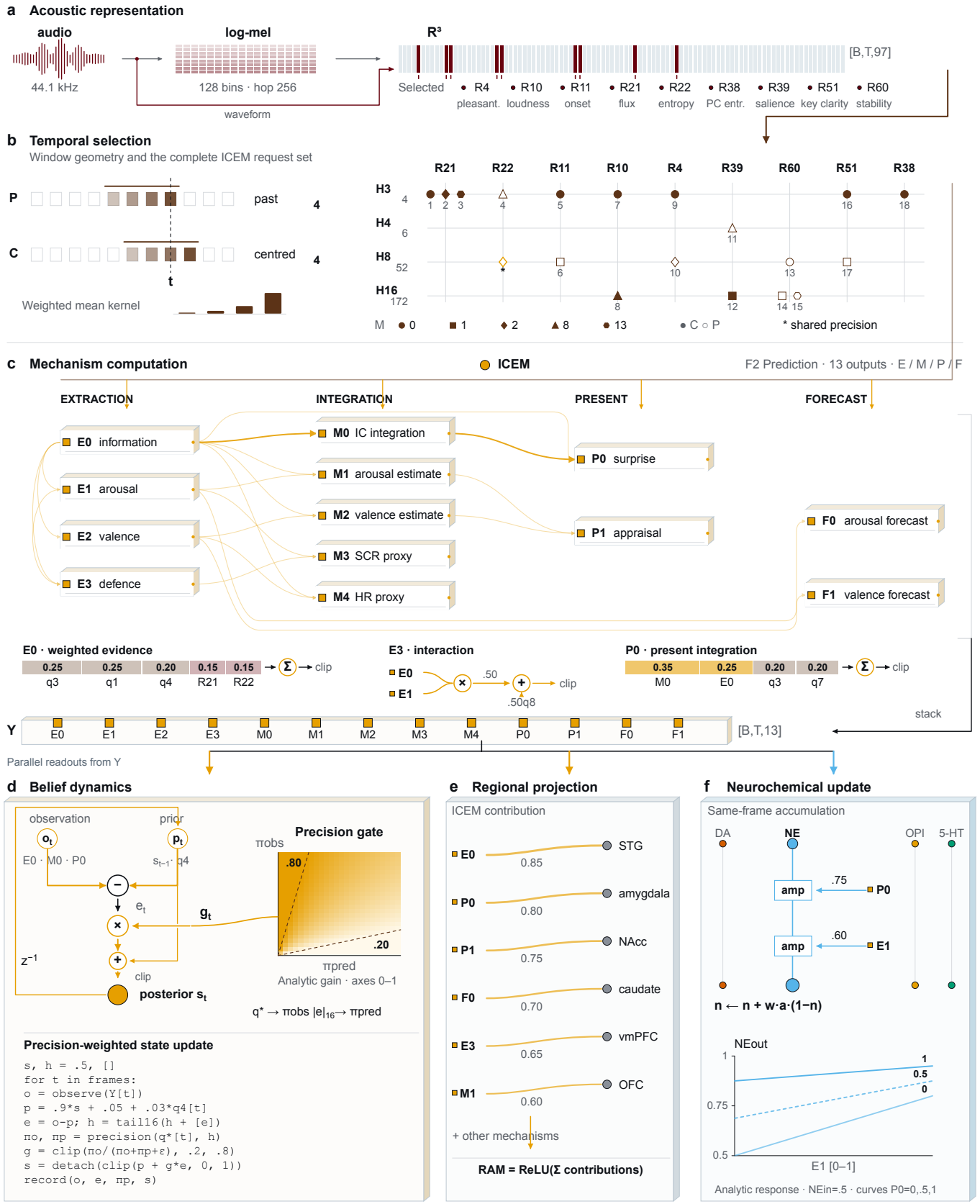


Figure 5. ICEM computation and parallel readouts. (a,b) Nine R³ channels and 18 temporal requests supply (c) 13 named outputs; insets expose weights and interactions. (d--f) Recursive belief, regional accumulation and ordered NE updates form parallel branches. The starred q^* precision input belongs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

Named mechanisms

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.

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 whether selected architectural dependencies can be reconstructed and revised in execution. Reconstruction recovers an output from saved inputs; changing one temporal request reveals which dimensions respond. Supplement S1 reports the recorded intervention and its protocol. Reconstruction records require controlled access.

An end-to-end trace

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

A targeted temporal intervention

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

Computational consequences

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

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

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

Computations before cognitive interpretations

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

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

Scope of the architectural evidence

The worked checks establish reconstruction and local revision for the inspected pathways. An acoustic quantity, a temporal request and a receiving calculation can be followed within one execution, including whether an available output is consumed. This makes a modelling assumption accessible to a controlled change. The examples do not establish that every connection is necessary, that all mechanisms interact beneficially, or that the full repertoire is needed for a given task. Those questions require comparisons across additional pathways and configurations.

Relation to existing frameworks

Reusable components and explicit computational graphs have precedents. Essentia supports audio-analysis networks (Bogdanov et al., 2013); the IPEM Toolbox provides perception-based music analysis; and AMT organises auditory models with demonstrations and reproduction experiments (Majdak et al., 2022). The PsyNeuLink environment connects psychological and neural model components, while ModECI specifies model structure and execution for exchange across environments (Gleeson et al., 2023). MI's proposed contribution lies in its particular organisation of named acoustic coordinates, explicit temporal requests and listening-related calculations with identifiable consumers. The worked intervention demonstrates one use of that organisation. Comparative development effort, efficiency and reuse across new hypotheses remain to be evaluated.

4. Empirical case studies

Empirical comparisons relate selected MI outputs to public observations: interval means, excerpt ratings, within-piece trajectories, chord rows and participant-level encoding. Direct associations assess correspondence; fitted readouts assess prediction. Each retains its protocol, observation scale and execution identity.

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.

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.

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.

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.

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).

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.

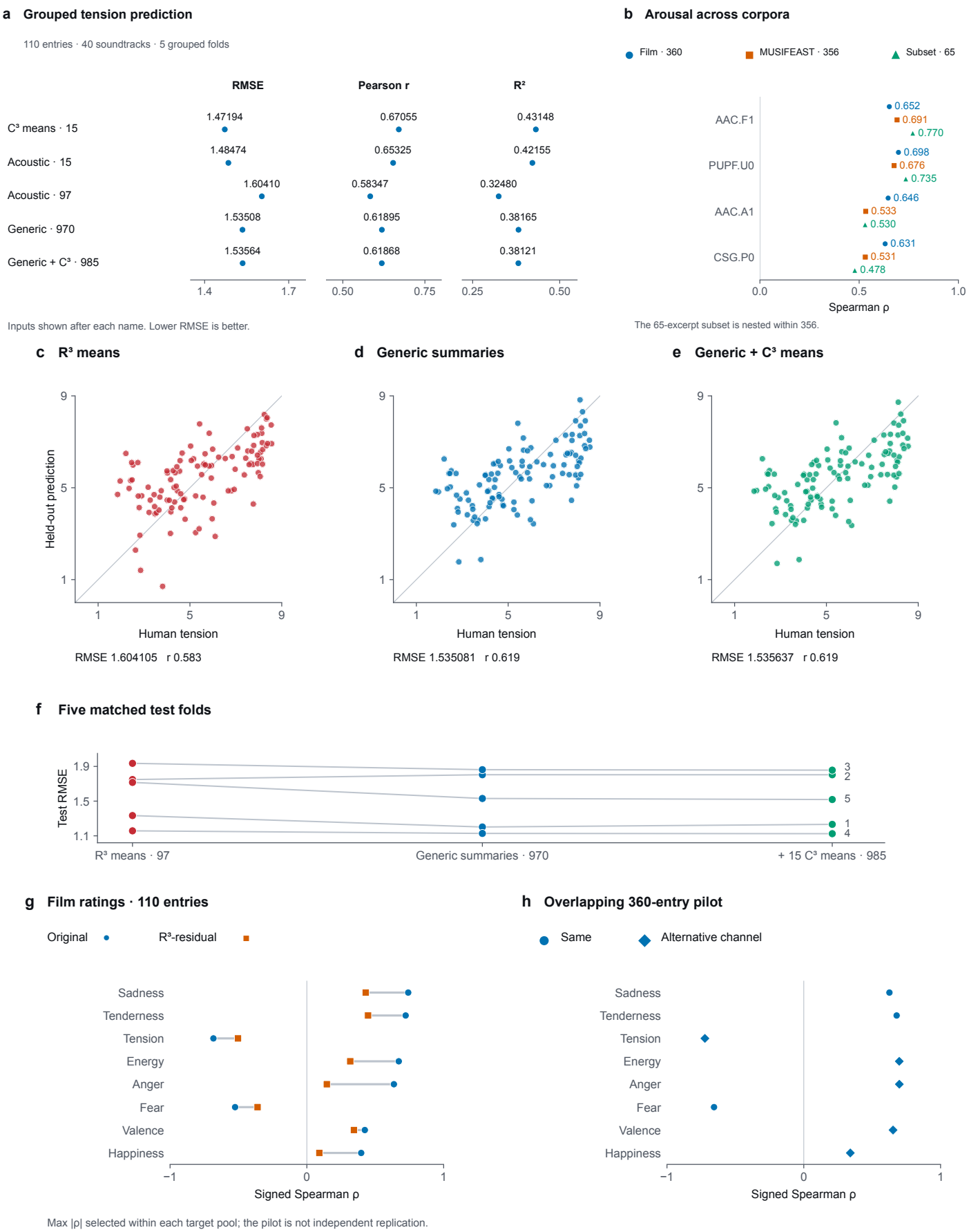


Figure 6. Excerpt-level prediction and cross-corpus arousal. (a) Five representations on 110 entries and five soundtrack-grouped folds; input counts follow each name. Acoustic 15 is the fixed comparator; generic 970 denotes acoustic/temporal descriptors, not canonical T³. RMSE is primary. (b) All four 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.

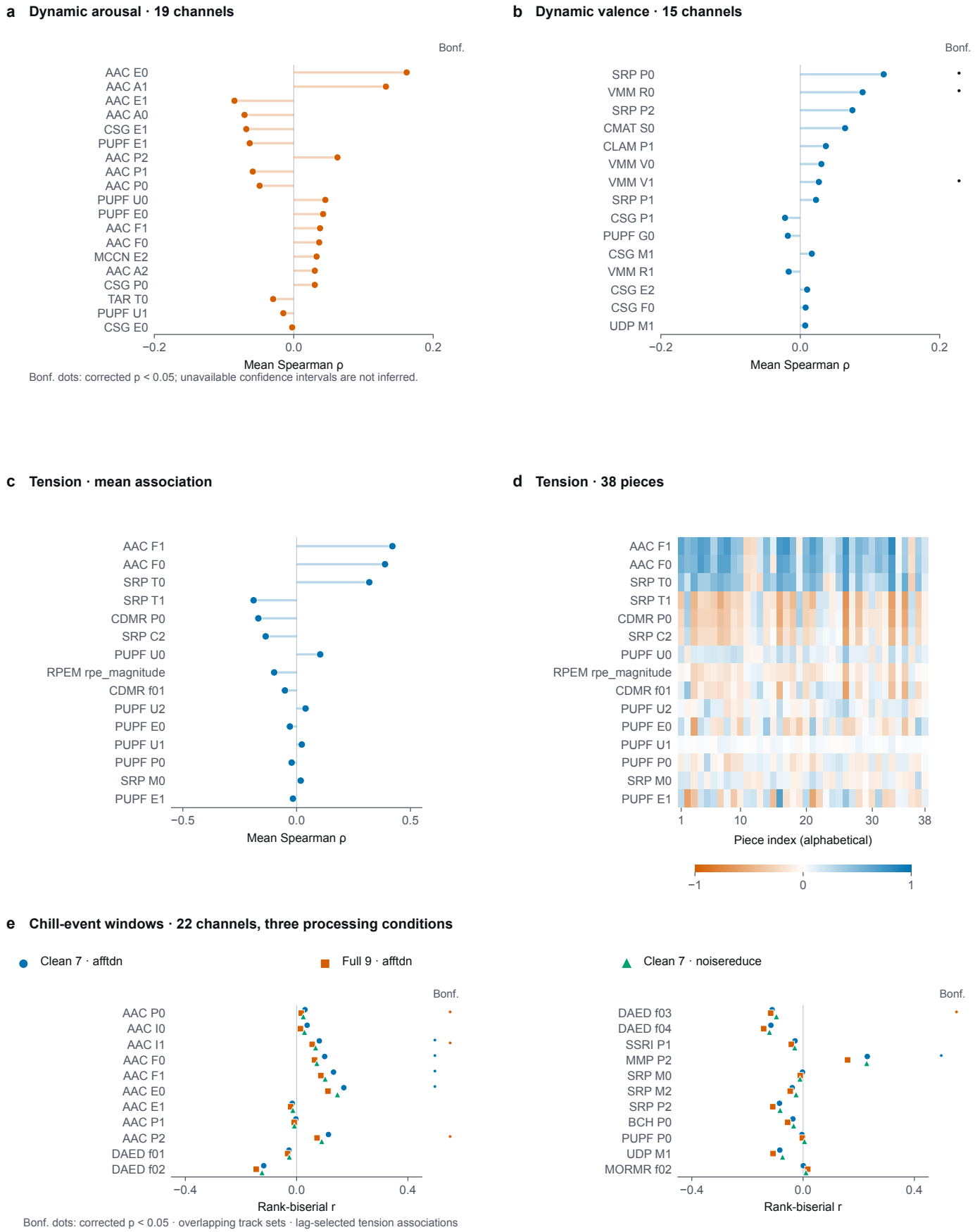
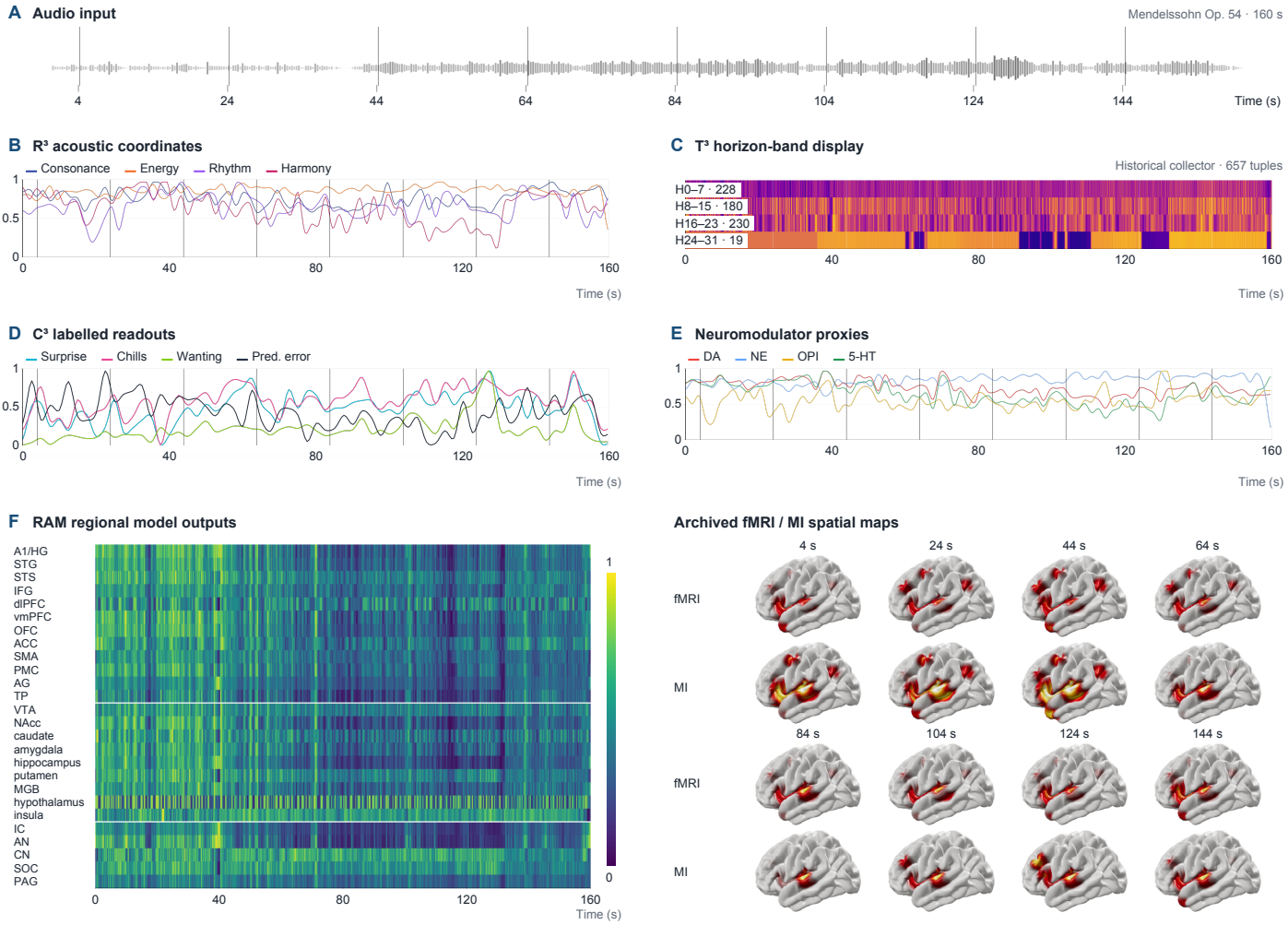


Figure 7. Time-resolved emotion, tension and reported chills. (a,b) All 19 arousal and 15 valence channels in the dynamic-emotion pilot, including negative and non-significant results. Points are saved Fisher-z means of lag-selected within-excerpt correlations. (c,d) All 15 mean tension associations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



G Acoustic channels across nine references

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

Spearman ρ: −1 0 1

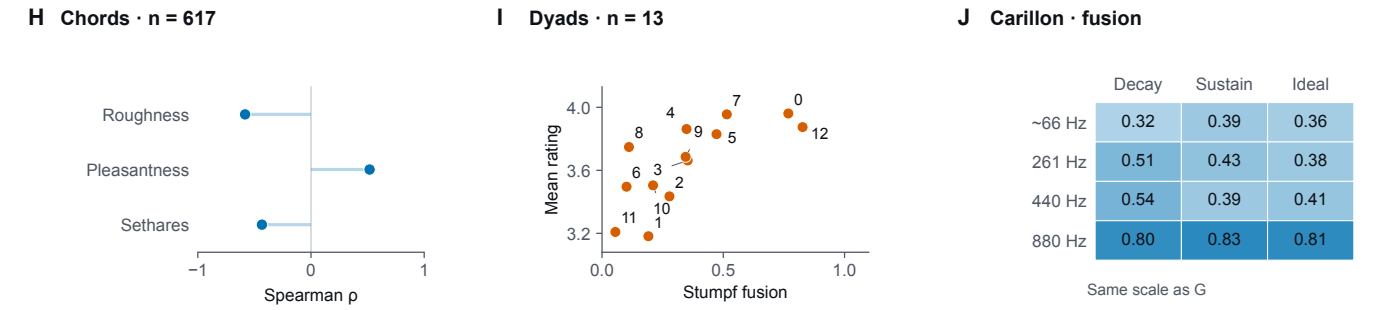
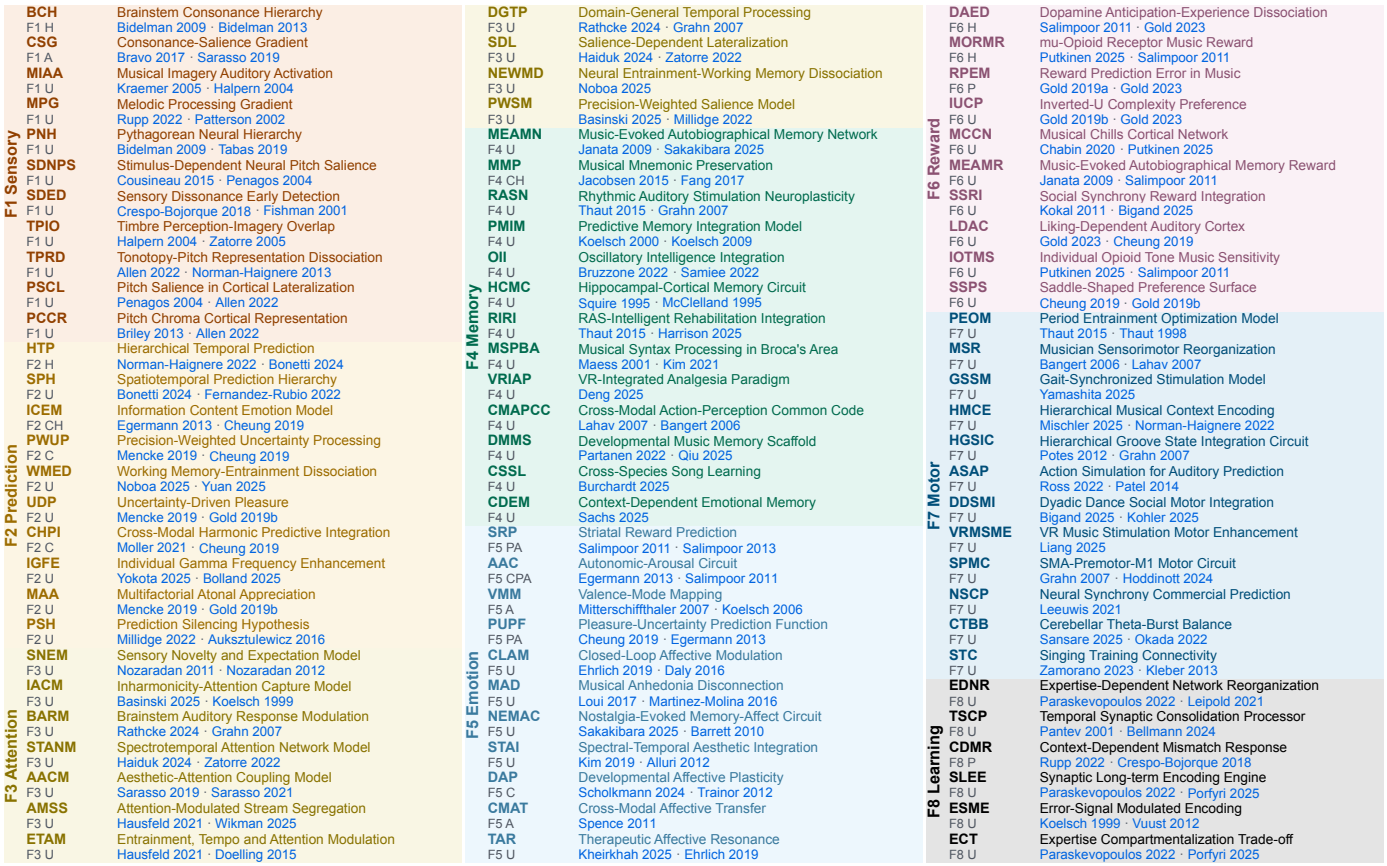


Figure 8. Acoustic representations and perceptual correspondence. (A–E) Archived 160-s Mendelssohn model displays retain source scaling; their 657-tuple collector differs from the current 644-address configuration. (F) RAM (left) and sixteen original brain views (right), pairing archived fMRI and MI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

Figure 9. 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 9 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.



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

5. Methods

Computational configuration and outputs

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

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

Worked execution and intervention

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

Grouped tension prediction

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

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

Arousal correspondence

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

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

Consonance comparisons

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

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

Public-data eligibility

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

Analysis scope

These evaluations are exploratory because corpus-guided development shaped the repertoire before comparison. The film corpus and tension pool were already known, and MUSIFEAST transfer followed an inspected pilot. Training-fold preprocessing and nested tuning constrain fitting within a specified model;

they cannot undo earlier representation selection (Cawley and Talbot, 2010). Bonferroni and BH corrections concern specified test families, leaving development exposure and effect-size superiority unresolved. Correlated channels and overlapping samples supply dependent evidence. Supplement protocols document negative results, historical execution identities and planning after data inspection. The narrow acoustic-comparator difference has no significance or equivalence test, and selected-output comparisons establish no necessity for all 84 mechanisms. Source definitions, analysis scripts and predictions are indexed in the accompanying record.

6. Discussion

The primary contribution of MI is an executable architecture for developing computational hypotheses about music listening. It makes acoustic inputs, temporal context and dependencies among calculations explicit, allowing selected assumptions to be revised and their consequences traced. The Swan Lake example demonstrates this use in a recorded execution: reconstruction recovers a selected output, and changing one temporal request produces effects within the declared route. Empirical case studies assess selected readouts under their own observation protocols. These evaluations address complementary questions about the architecture's operation and the use of its outputs.

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 present demonstrations establish this use for selected pathways. Broader usefulness requires showing that the same organisation supports additional hypotheses and controlled comparisons.

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 limit the predictive case for the larger repertoire. They leave the utility of explicit specification and revision as a separate question. Efficiency, upstream cost and the time required to develop or revise a model remain to be measured.

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).

Research use and the next tests

Three comparisons would extend the architectural evaluation: replace a selected composite with its acoustic constituents under the same observation protocol; revise temporal context across representative direct, shared and inter-mechanism pathways; and introduce measured listener history where an interpretation requires it. Restricting a forecast pathway to past information must include upstream normalization as well as its T^3 window. These are proposed tests. Their value would lie in distinguishing computational alternatives while keeping the remaining assumptions explicit.

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.

Scientific Supplement

Implementation records: controlled access

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.