Individual expectations and adaptive prediction of musical pleasure

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

Abstract. This protocol separates learning by a listener from learning about that listener. Twenty familiar classical works initialize a constrained personal model; controlled tasks test an uncertainty–surprise interaction and updating after randomized changes in continuation probabilities. A common 24-excerpt transfer block separates selection-policy value from predictive accuracy. Matched comparisons isolate personal references, earlier expectation reports and MI representations beyond an equally informed acoustic learner. Three primary claims share familywise error control. A 60-registration pilot within a 1,000-registration cap qualifies feasibility, measurement, model recovery and power before confirmation. No participant results or achieved power are reported.

Prospective Stage 1 preparation. Figures are original analytic or conceptual schematics, not participant observations.

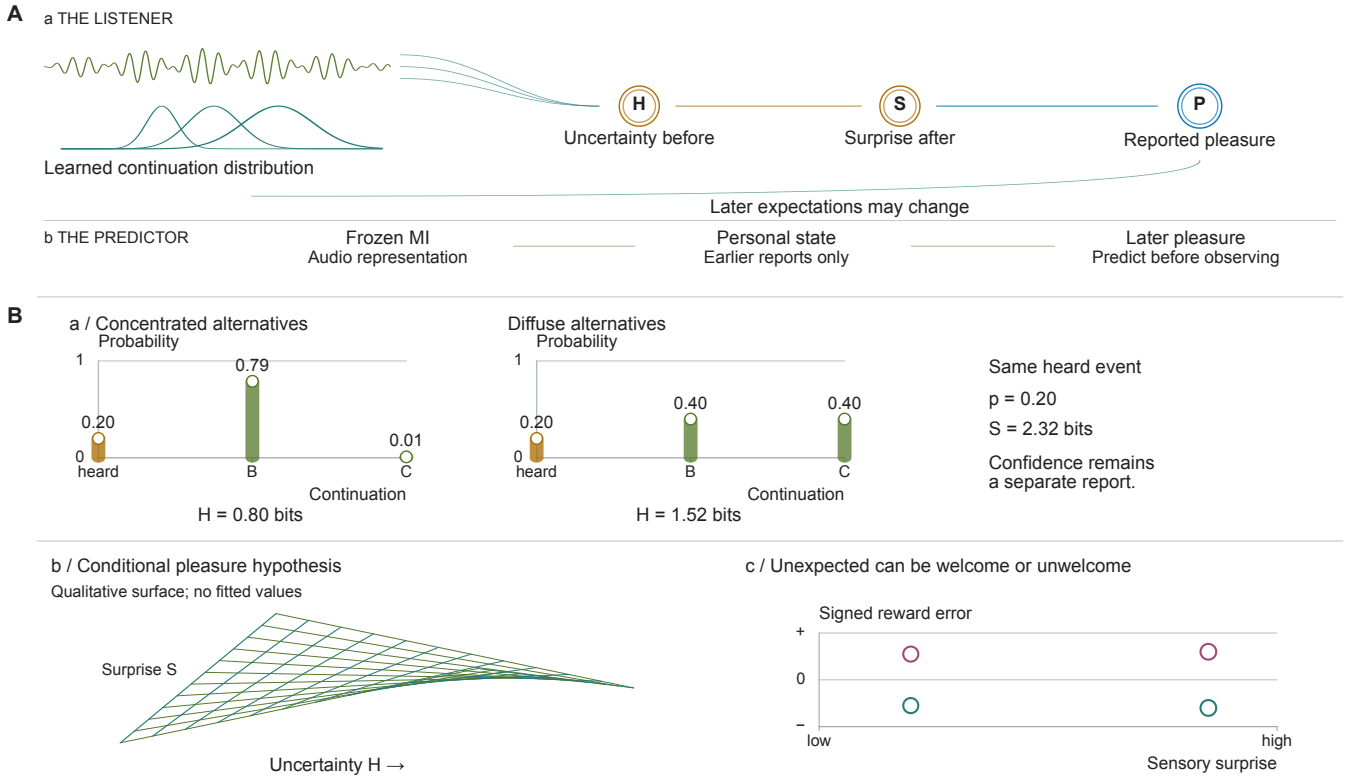


Fig. 1. Expectation and prediction. A, Listener updating and predictor updating are distinct. **B,** Exact distributions yield the same surprise (2.32 bits) but different entropy (0.80/1.52 bits); the pleasure surface is qualitative, neither a fit to Cheung et al.[2] nor MI's PUPF equation. Its normalized axes do not enumerate all feasible distributions. Sensory surprise can accompany positive or negative value error.

Scientific question and operational definitions

Predictive accounts place listening history between sound and pleasure[1]. Evidence for conditional uncertainty–surprise effects[2], intermediate predictability preferences[3] and individual predictive preferences[22] motivates three separable questions: whether uncertainty moderates the surprise–pleasure association; whether randomized exposure changes subsequent expectations; and which information improves prediction of later pleasure. These are, respectively, an observational association within a controlled task, a randomized behavioural contrast and an out-of-sample computational comparison. The contribution is their coordinated test, not a first computational theory of musical pleasure.

For a listener i on trial t, $q_{it}(k)$ is the reported probability of three demonstrated continuations and e_t the heard outcome:

$$H_{it} = -\sum_{k=1}^{3} q_{it}(k)\log_2 q_{it}(k), \qquad S_{it} = -\log_2 q_{it}(e_t).$$

These task-specific quantities are distinct from confidence and from a person's full musical expectations. Expected enjoyment V_{it} and experienced pleasure Y_{it} share 0–100 anchors; signed value error is $\delta_{it} = Y_{it} - V_{it}$, whereas prediction residual is $r_{it} = Y_{it} - \widehat{Y}_{it}$. Neither is sensory surprise. Lower entropy can reflect overconfidence; learning requires later changes in expectations, choices or accuracy.

Arousal associated with unexpectedness[4], perceived valence/arousal predicted from acoustics[23], anticipation, reward value and peak pleasure[5, 6], and auditory–motor learning and recognition[24] are different endpoints. Reward-learning[7] and naturalistic expectancy–liking evidence[8] motivate tests without establishing a dopamine mechanism in this browser study. Nonlinear complexity preferences, familiarity, acoustic salience and rating habits remain competing accounts. Behavioural support and MI agreement are assessed independently.

Participants, materials and session structure

Recruitment and accounting. The cap counts 1,000 distinct, verified, minimally eligible registrations, not completed datasets. Verification, age eligibility and initial screening count a registration once; study consent is a separate versioned event. Withdrawals/incomplete baselines remain in the flow and do not reopen slots. Report invitations, verification, screening, consent, completion of 20 references, readiness, arm completion, transfer and analysable endpoints. The first 60 registrants form a feasibility/parameter-setting pilot excluded from confirmation; the remaining cap is 940 before attrition, without an assumed power guarantee.

Adults aged at least 18 must be able to consent in the study language and use supported headphones and a desktop/laptop. Record musical training years, current practice, classical listening, language, age band, country, self-reported hearing difficulty and assistive hearing use, allowing missing/prefer-not-to-answer where appropriate. Sample across training and familiarity; do not infer them from demographic stereotypes or infer diagnoses. Start with one validated language; translations require cognitive interviewing and counterbalanced validation before pooling.

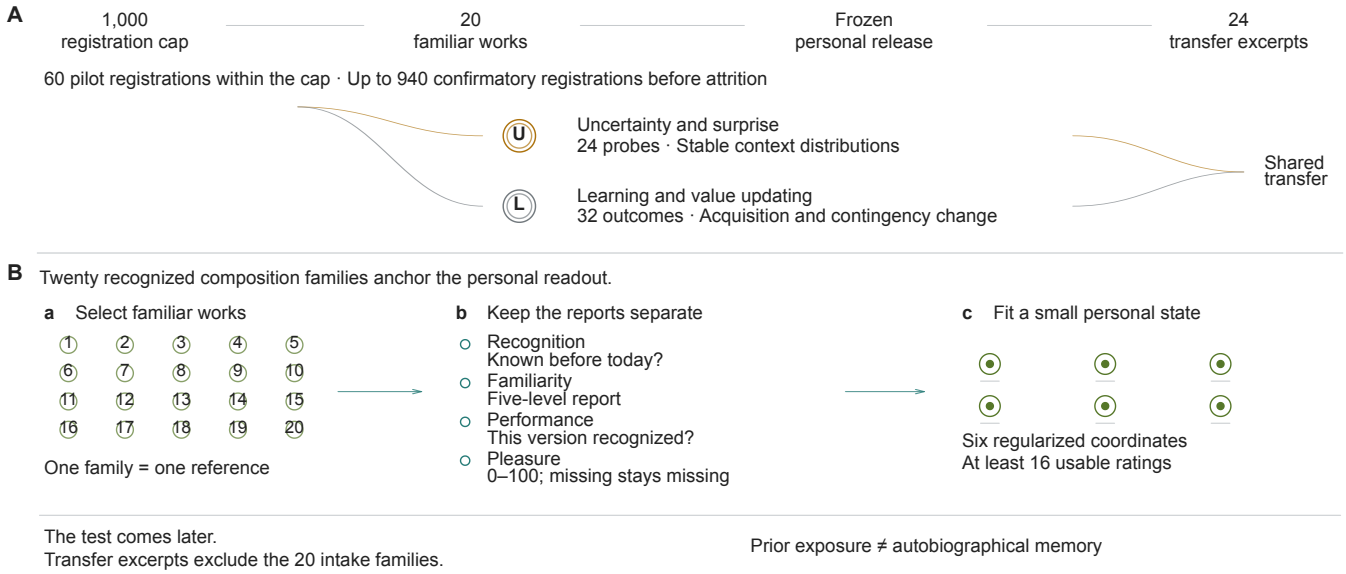


Fig. 2. Cohort and intake. A, Proposed allocations: 60 pilot registrations within the cap, at most 940 remaining before readiness failures or attrition; equal U/L assignment precedes shared transfer and four performance pairs. **B,** Twenty distinct, previously recognized families and at least 16 usable pleasure ratings initialize six regularized coordinates. Transfer excludes the intake families. Selection is not a measurement of autobiography or general musical ability.

Catalogue and rights. Track composition family, movement, arrangement, performance, recording master, excerpt, passage, duration, instrumentation, performer attribution and study exposure. Re-encodings are not new performances; movements cannot fill separate intake positions. Candidate sources include individually cleared classical recordings and MAESTRO[18]. Each exact recording/derivative needs a retained licence covering use, territory, redistribution and attribution; neither composition age nor dataset access supplies it. Freeze the catalogue after audit; no YouTube downloading is proposed. Full transfer coverage need not provide three performances per family; only the paired-performance subset does.

Intake. In a separate visit, a searchable catalogue supplies 12-second reference previews. Exactly 20 distinct families must be recognized from before today, including disliked works. Collect prior recognition, five-level familiarity, recognition of this performance and preview pleasure (0–100, separate unable-to-rate option). Do not substitute onboarding repetition for prior familiarity, infer streaming history or require a streaming account. At least 16 usable pleasure ratings are needed for calibration; incomplete intake can be saved but requires rating resolution before readiness. Unsupported coefficients shrink to the pilot-trained population prior. Pilot outcomes include search burden, completion proportion and background-related selection; the 20-work requirement narrows representativeness.

Experimental visit. Consent, browser/audio checks and intake precede preparation of an immutable release. The visit starts with comfortable-volume adjustment, headphone screening[19] and comprehension trials distinguishing likelihood, confidence, surprise and enjoyment. Persistent equipment/comprehension failure is recorded before randomization; screening does not establish calibrated acoustics. One mechanistic arm, a break, common transfer and performance pairs follow. The proposed combined intake/experiment duration is 55–75 minutes, to be replaced by pilot timings; pauses between blocks and withdrawal are allowed. Finalize institution, payment, recruitment, consent and contacts before invitations; rewards never depend on liking or agreement with a hypothesis.

Transfer and performance controls. Use 24 previously unpresented 25-second excerpts excluding intake families, with six each selected by personal MI, an equally calibrated acoustic model, matched random sampling and MI with another person's profile. Outside-study familiarity remains possible and is reported after playback. Use standardized fades and fixed loudness normalization preserving internal dynamics; reject clipping, noise, unmatched truncation or inconsistent channels. Retain loudness, roughness, tempo, onset density, tonal stability and instrumentation as controls. Four additional pairs use two previously unheard performances beyond the intake reference, matched on passage, instrumentation and approximate position; counterbalance B/C order, collect both pleasure ratings and forced preference with an indifferent option. Arrangement changes require a separate study.

Controlled expectation and learning tasks

Both arms use cleared or newly rendered classical-motif continuations, with identical target endings available across contexts. Record intended distributions, seeded realizations, actual delivered exposures and counterbalanced outcome identities separately; schedules are not selected by later pleasure, and no online trial is inserted to force an effect. Before each outcome, allocate 100 probability points among three demonstrated audible alternatives, then report confidence and expected enjoyment. No notation knowledge is required. Lock priors, replay the context to restore a defined auditory lead-in, and record delay/replay; surprise and pleasure follow the ending. Apply the same reporting order in both arms.

Arm U: conditional pleasure. Two visually labelled contexts have distributions (.20, .75, .05) and (.20, .40, .40), with outcome identities permuted across listeners. Twenty-four training outcomes (12/context) precede 24 probes using non-overlapping sequence realizations, balanced context order, six-second contexts and three-second endings. Collect arousal on eight prespecified probes balanced by context/outcome class. Preserve zero probability reports; floor at .01 and renormalize only for logarithmic scoring, with .005/.02 sensitivities. Qualify simplex comprehension, empirical entropy–surprise support and floor/ceiling behaviour in the pilot; freeze waveform joins and timing afterwards. Equal scheduled target probability does not imply equal subjective surprise, and subjective entropy/surprise are not independently randomized.

U estimand and model. Fit 0–100 pleasure with listener-centred/scaled H, S, HS, H^2, S^2, context, order, reference familiarity and acoustic controls; include participant and continuation-family intercepts and identifiable participant H/S slopes. For singular fits, first remove random-effect correlations, then the least-supported slope, retaining both intercepts. The directional interaction predicts a negative coefficient, accompanied by positive surprise slope at the pooled 25th percentile and negative slope at the 75th percentile of centred pre-event entropy, defined without outcomes. Report joint uncertainty; interaction alone does not establish both signs. Use a bounded-outcome sensitivity model. Shared q can induce correlated measurement error; check floors, reliability and occupied support, distinguish nonlinear-main-effect

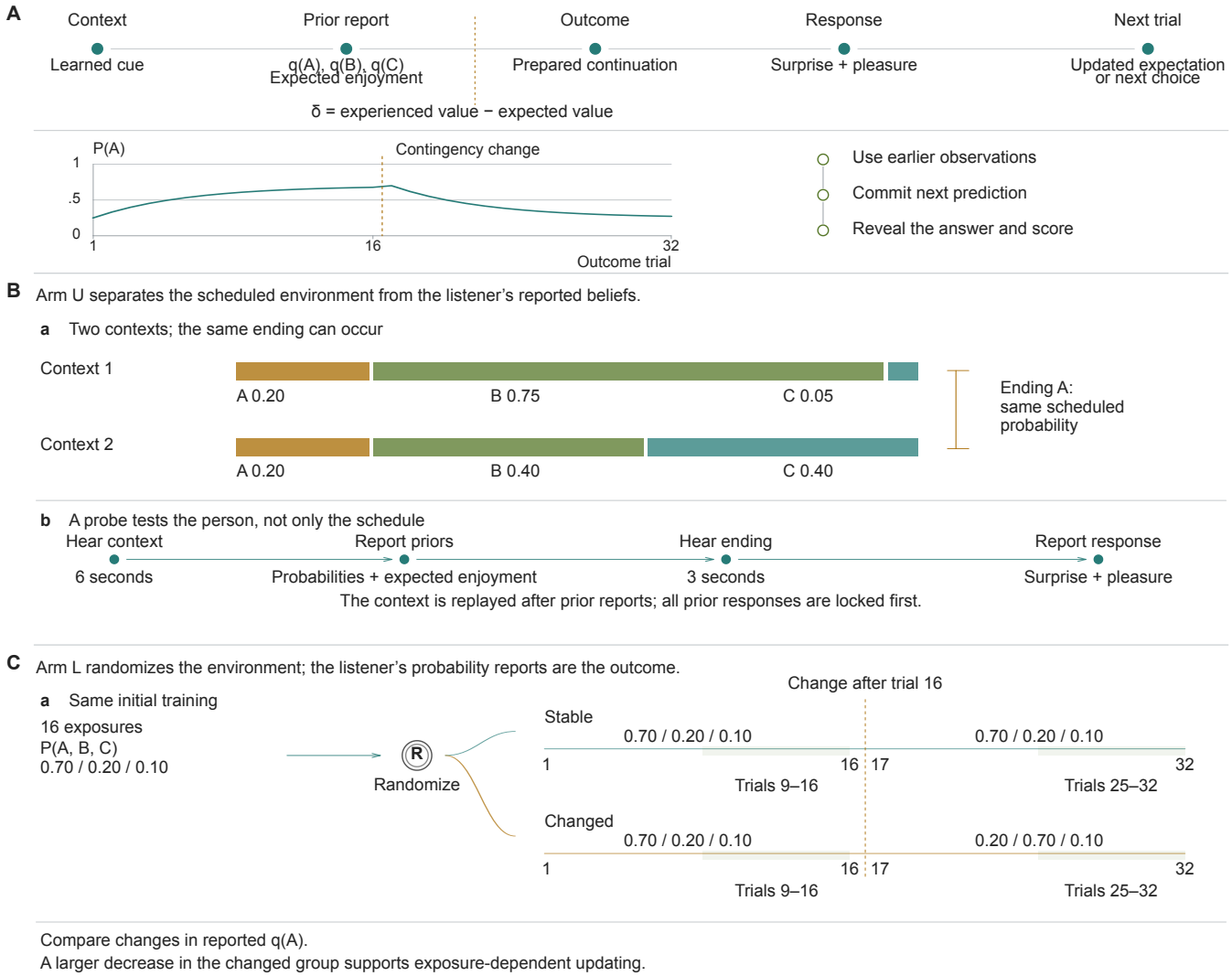


Fig. 3. Controlled tasks. A, Lock priors before the outcome; measure responses afterwards. The adaptation curve is analytic. **B,** U separates scheduled distributions from reported beliefs; an ending has scheduled probability 0.20 in both contexts. **C,** L contrasts stable with changed distributions after probe 16. Displayed probabilities describe schedules, not observed learning. Replay preserves the fixed listening sequence and never makes collection retrospectively adaptive.

explanations by recovery, and do not retune quantiles, lags or polynomial order on confirmation.

Arm L: exposure-dependent updating. Sixteen training exposures from $(.70, .20, .10)$ precede 32 probes. Randomize half to switch to $(.20, .70, .10)$ after probe 16; half remain stable. Independently counterbalance outcome identities. For initially dominant A, let $\Delta\bar{q}_i(A)$ compare probes 25–32 with 9–16. The primary assignment effect is $\tau_L = E[\Delta\bar{q}_i(A)|Z_i = 1] - E[\Delta\bar{q}_i(A)|Z_i = 0]$. Fit condition-by-phase with participant intercepts, baseline probability, counterbalanced identity and realized-exposure imbalance; never treat post-assignment pleasure as a pretreatment covariate. A larger decrease under change supports updating regardless of pleasure. Report arm-specific completion, baseline predictors of missingness and all-available-data sensitivities. Next-outcome log score, adaptation rate and signed value learning are secondary; short sequences require hierarchical uncertainty, not diagnostic learning-rate rankings.

Personal models, fixed releases and policy value

Release and replay. Compute each versioned audio representation once. Before confirmatory recruitment freeze global mappings, dimension reductions, module selection, learner priors/updates and tuning grids. Personal preparation uses 20 reference ratings and declared background to save the full schedule, fixed predictions, policy assignments, seed and content hashes. Collection requires no MI, GPU or LLM inference. Chronological offline replay receives only intake, the completed mechanistic task and earlier transfer responses: save state, commit prediction, then reveal the current target. Known-stimulus prediction may use full prepared audio; it does not establish the listener's time-local anticipation. Fixed-release and replay commitments remain distinct records.

Model specification. With acoustic, temporal and mechanism features x_j^A, x_j^T, x_j^M, intake B_i, state z_{it} and observed outcomes o_{it}, use $z_{i0} = G(B_i;\theta)$, $z_{i,t+1} = U(z_{it}, o_{it};\theta)$ and $\widehat{o}_{it} = F(x_j^A, x_j^T, x_j^M, z_{it};\theta)$. These specify information flow; constrained functions, priors and global θ must be instantiated/frozen at pilot exit, not assumed already trained. Mask uncollected outcomes. The primary linear readout uses six standardized coordinates: two pilot-catalogue principal components from each feature group, versus six acoustic components in the acoustic baseline. A population ridge model supplies coefficient priors; personal fitting minimizes reference squared error plus shrinkage, with penalty $\{.1, 1, 10, 100\}$ selected by participant/composition-grouped pilot validation. Retain all transforms, priors, coefficients and software hashes. Nonlinear and richer state-space learners are secondary, with matched tuning budgets.

The pleasure-only learner cumulatively updates after each eligible earlier rating. The expectation-informed version adds a global pilot-fitted adjustment using earlier mean entropy/surprise, expected enjoyment, confidence calibration and change in continuation accuracy. Mechanistic summaries initialize transfer; subsequent surprise/pleasure update only remeasured features, without inventing excerpt continuation probabilities or adding personal fit coefficients. Acoustic comparators receive identical reports, background, timing and fitting resources. Replay and parameter-recovery checks precede confirmation; shuffled-person states and expectation-report permutations test specificity.

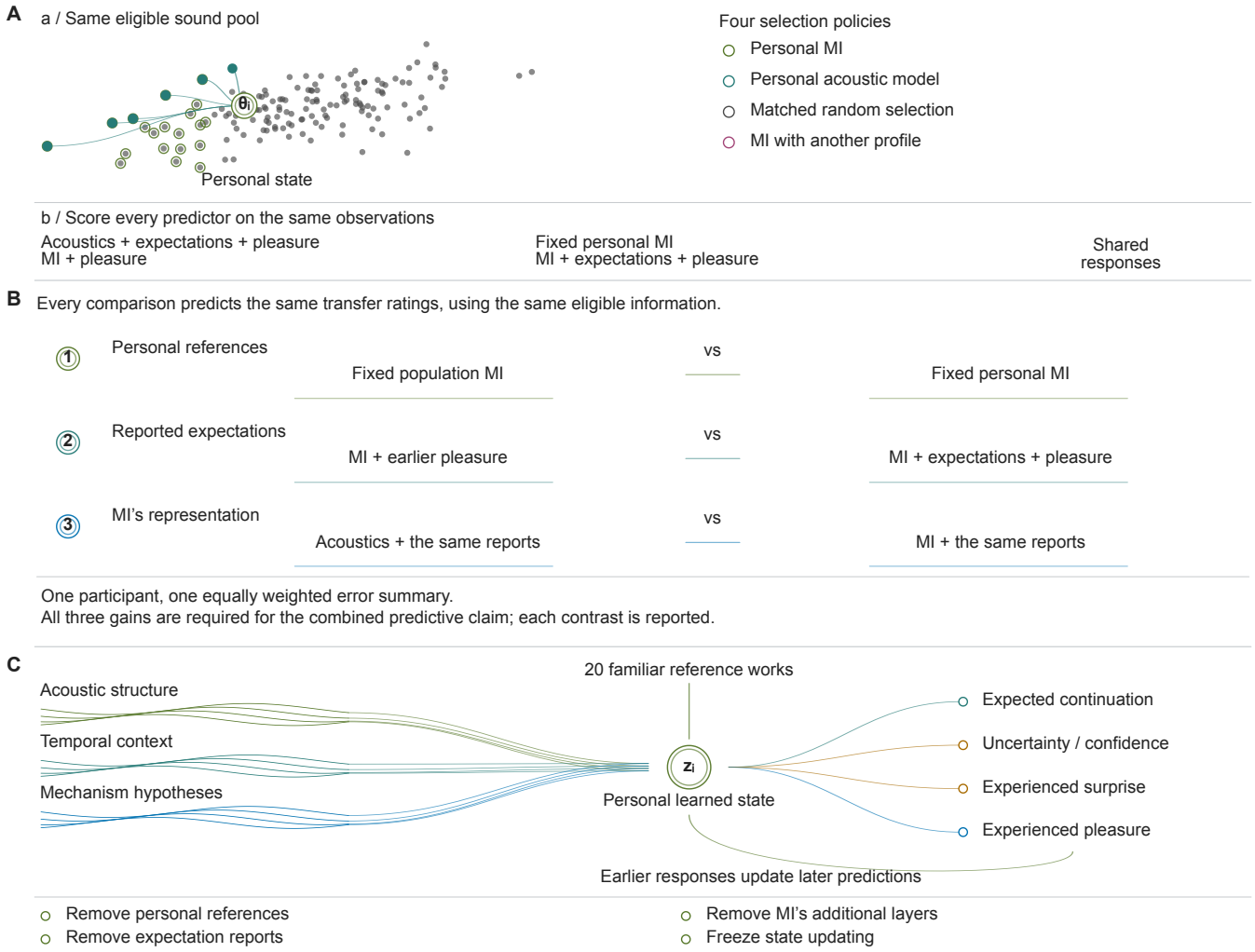


Fig. 4. Personal prediction. A, Four policies share an eligible pool; synthetic display coordinates are not fitted embeddings. Every predictor scores the same observations. **B,** Three matched predictive increments must all improve for the composite claim; individual estimates remain visible. **C,** References and earlier responses inform a personal state; matched ablations remove history, layers, expectation or updating. Connections describe computations, not neural pathways.

Predictive target. On identical eligible trials $\mathcal{T}_i$, define $L_i(m) = |\mathcal{T}_i|^{-1} \sum_{t \in \mathcal{T}_i} |Y_{it} - \widehat{Y}_{it}^{(m)}|$ and $\Delta_{a,b} = E_i[L_i(a) - L_i(b)]$. Positive differences favour enrichment; each listener has equal weight. Require fixed personal MI over fixed population MI, expectation-plus-pleasure MI over pleasure-only MI, and expectation-informed MI over an equally updated acoustic learner. The composite intersection–union p is the maximum of the three one-sided p values; report each estimate/interval. Holm correction at $\alpha = .05$ covers this composite, U and L. Clip every prediction to 0–100 using the same frozen rule; retain personal-mean/intercept and repetition/order controls.

Selection policy. Randomly partition the same eligible pool into disjoint policy candidate sets before ranking; rotate membership across participants. Conceal labels and interleave slots in balanced six-trial blocks, with six per policy overall. Random selection matches instrumentation, duration and broad familiarity, not MI score; MI/acoustic policies share profile and ranking budget. Match swapped profiles by training and catalogue coverage without identities. Unfillable quotas block release. Policy value concerns complete policies in the released pool; it is neither fixed-audio module causation nor an unrestricted catalogue advantage. Alternative-population reweighting needs known selection probabilities and support. Keep policy value, predictive accuracy and mechanism support separate.

Mechanism audit and secondary explanations

The executable MI architecture[9] provides candidate computations whose names do not establish construct or biological validity. The nine-module audit preserves the fixed implementation even if behavioural results disagree; all tests outside the three primary claims remain secondary or exploratory.

HTP and ICEM. Hierarchical Temporal Prediction draws motivation from auditory integration and learned sequence prediction[10, 11], but its stability, pitch, timbre and onset inputs are not a listener's probability distribution. Information Content Emotion Model uses acoustic variation, not $-\log_2 q(e)$. Arm U tests correspondence with elicited expectation/surprise beyond acoustic change; separately sampled arousal supplies no cardiac or skin-conductance validation.

PWUP and UDP. Precision-Weighted Uncertainty Processing scales raw error as $E(1 - .5g)$ using its gate g; its name does not determine empirical direction. Uncertainty-dependent neural deviance responses[12] motivate a separate test of this gate. Uncertainty-Driven Pleasure contains an upstream uncertainty/confirmation branch and a distinct acoustic pleasure index; test their increments separately. Theory about uncertain-music pleasure[13] motivates, without validating, either computation.

SRP and PUPF. Striatal Reward Prediction provides audio-derived wanting, liking and tension proxies; tension research[14] does not equate tension, relief and liking. Pleasure-Uncertainty Prediction Function combines acoustic proxies and pleasantness; it is not a fitted copy of the empirical uncertainty–surprise surface. Retain its frozen equation even when it fails.

DAED, RPEM and WMED. Dopamine Anticipation-Experience Dissociation and Reward Prediction Error in Music are acoustic computations; centred RPEM is not reported $Y - V$ and does not itself update a personal learner. Working Memory-Entrainment Dissociation is motivated by separable timing/memory evidence[15, 16]; the core task measures neither working-memory capacity nor neural entrainment. Without dedicated assays, assess these outputs only against actually observed targets.

Joint model and ablations. A secondary explanatory model combines probability reports, unexpectedness and pleasure, with participant/composition effects and performance nested within composition. Use a logistic-normal model for the probability simplex, bounded affect models, and distinct expected enjoyment,

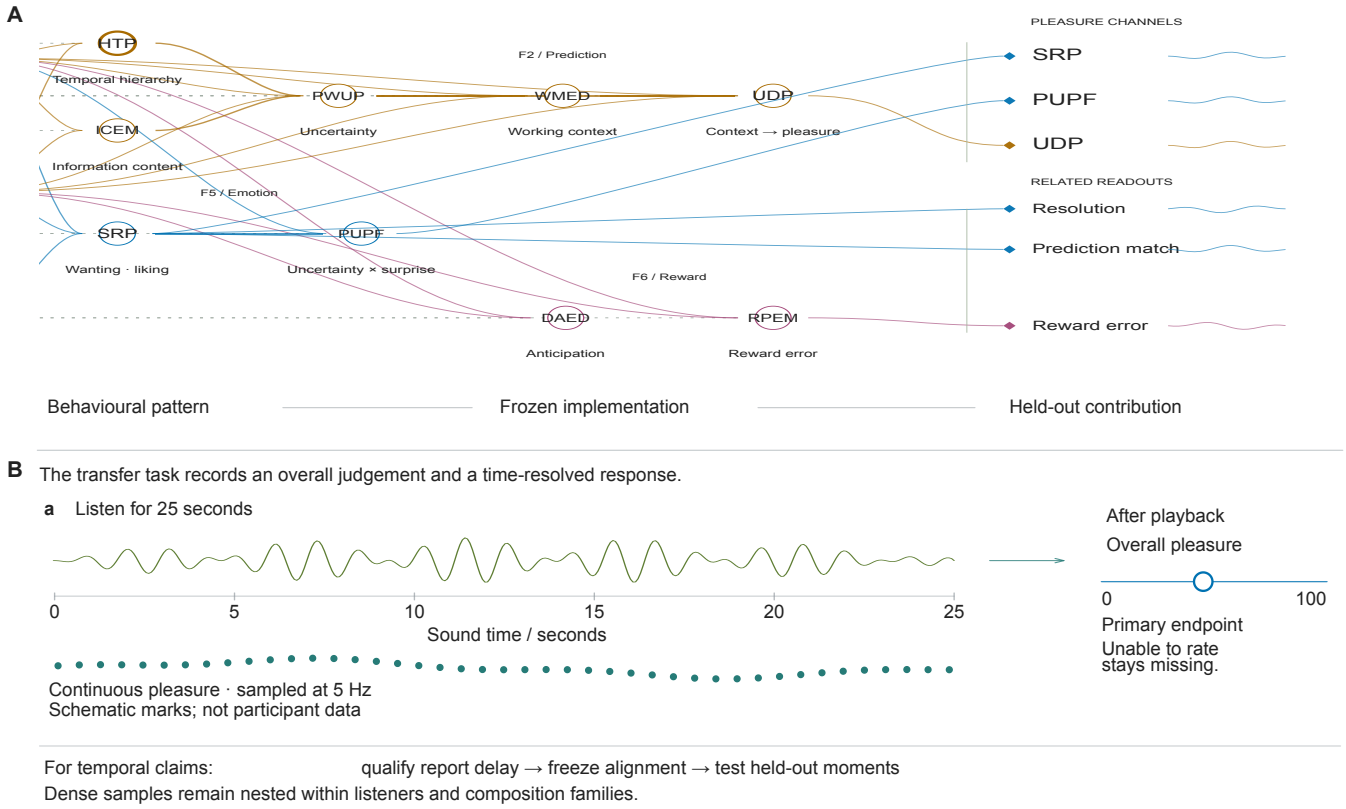


Fig. 5. Mechanisms and measurement. A, The original nine-module topology and function colours are retained: six within-set dependencies plus incoming acoustic/temporal inputs and outgoing readouts. RPEM/DAED gain no invented upstream mechanisms. **B,** A 5-Hz trace accompanies each 25-second excerpt; the immediate overall rating is primary. All traces and slider positions are schematic. Dense samples remain nested within listeners and composition families.

surprise and signed error; H/S from one distribution are not independent observations. Include acoustic controls, temporal/mechanism features, response dependence and prespecified training/familiarity moderation. Equal-capacity/tuning ablations compare acoustic, acoustic-plus-temporal and full layers; remove calibration, freeze updating or omit probability/expected-value reports while retaining pleasure. Correlated outputs permit incremental/overlapping information, not unique causal attribution. Reference-dose analysis uses 0, 5, 10 and 20 families with ten preregistered subset orderings on the same test excerpts; report missing ratings and effective counts.

Secondary value and timing assays. Compare $V_{t+1} = (1-\alpha)V_t + \alpha Y_t$ with fixed value, repetition, unsigned surprise and experienced-reward alternatives by next-trial enjoyment/choice prediction under identical prior information. Correlating $Y_t - V_t$ with $V_{t+1} - V_t$ is algebraically coupled; regularize rates and choose hyperparameters only in the pilot. Separately consented extensions compare resolved/delayed/withheld endings with pre-ending wanting/enjoyment and post-ending relief/liking; separate anticipation and experience response blocks; and rhythmic regularity crossed with two-/four-item pattern memory, measuring tapping error and recognition separately. Their absence limits SRP/DAED/WMED interpretations.

During transfer, sample continuous pleasure at 5 Hz; afterwards record overall pleasure (0–100, unable-to-rate distinct), unexpectedness and recognition. The trace maximum is not the primary endpoint. Declare candidate pleasurable moments from prepared signals, compare matched moments within/across excerpts, and freeze pilot-qualified reporting lag/smoothing; never optimize lag or peak on test traces. Record audio, browser-monotonic and server-receipt clocks separately and qualify latency/variability on representative hardware[20].

Analysis, feasibility and reporting

Dependence and uncertainty. Estimate paired predictive differences with a crossed participant-by-composition bootstrap (5,000 replicates), retaining family assignments across occurrences. A predeclared cluster wild-bootstrap sensitivity handles sparse work coverage; verify the resampling procedure's error control in power simulations. Mean squared error, calibration slope and signed bias are secondary. Estimate policy pleasure by mixed models with participant/work effects, order and prespecified familiarity controls. Apply false-discovery-rate control within secondary mechanism, policy and temporal families; exploratory findings require independent confirmation. Dense traces, repeated works and many listeners support different effective sample sizes.

Recovery and meaningful effects. Simulate dependence among participants and compositions, bounded ratings, incomplete baselines, actual allocation, clustering, attrition and differential dropout. Nulls include acoustic-only pleasure, scale differences, nonlinear main effects without interaction and calibrated expectation without hedonic contribution; positive controls span relevant effects/nuisance variation. Each analysis must recover its own target without falsely recovering another[25]. Retain inputs, seeds, Monte Carlo uncertainty and failed scenarios. Do not rank individual coefficients as psychological traits without recovery and independent reliability; nonsignificance is not equivalence.

Pilot estimates must establish intake/comprehension feasibility, H/S coverage, variance, dependence, missingness, balance and timing. Target at least 90% power at the worst Holm threshold, with Type I error checked under null/misspecified scenarios. Initial meaningful-effect targets are standardized interaction magnitude .10, changed-versus-stable probability difference .10 and one point less absolute prediction error on the 0–100 scale for each predictive contrast. These are planning thresholds, not literature estimates; justify/amend by reliability and substantive interpretation before confirmation, never for favourable significance. Simulate 300, 500, 700 and 900 analysable participants. Revise scope/recruitment if attainable precision is insufficient; retain the pilot-exit amendment and executable simulations. The registration cap establishes neither effective sample size nor achieved power.

Missingness and exclusions. Prespecify consent, identity-duplication, technical-delivery and task-quality rules; never exclude low/constant pleasure or disagreement with MI. Primary prediction requires at least 18/24 valid transfer ratings; arm summaries require at least 75% valid probes and non-empty data in both phases/contexts. Report all-valid-observation sensitivities and intake predictors of completion. Invalid trials include failed audio start, stalls, target-period hidden pages or outcomes inconsistent with release. Late upload alone is acceptable when local timing and

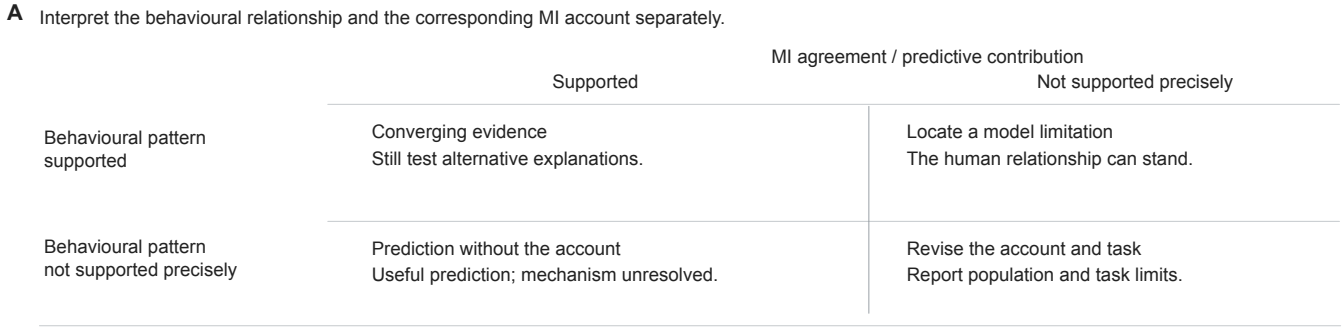


Fig. 6. Interpretation. Behavioural support and the corresponding MI claim are evaluated separately. Convergence still requires alternatives; behavioural support with MI failure locates a model limitation; prediction without the proposed pattern leaves its explanation unresolved. Precise incompatibility narrows the operational account; wide intervals are inconclusive. The quadrants are possible outcomes, not findings.

verified receipt survive. Replay is a new exposure, excluded from primary first-exposure analyses and retained for sensitivity; it never silently replaces the first response. Do not forward-fill reports or impute primary outcomes. Baseline-covariate imputation, if used, is pilot-frozen with missingness indicators.

Inference and scope. Behavioural learning and computational personalization are coequal aims. An uncertainty–surprise pattern with no MI increment limits the representation/readout; prediction without that pattern supports utility with unresolved mechanism; precise absence of both challenges this operational account. Controlled three-option expectations and naturalistic transfer need not agree; failed transfer is a substantive limit. Familiarity, acoustic salience, reporting reactivity and response style remain plausible despite acoustic, personal-mean and within-work controls. The catalogue's 20-family requirement selects listeners, and diverse styles/cultures require new coverage/tasks/validation. Results cannot identify neural mediators, anatomy, neurotransmitters, autobiography or general musical ability from behavioural ratings or module labels. Broad predictive-processing theory is not refuted by one task-specific failure.

Ethics, access and status. Consent must specify tasks, burden, stored information, contact separation, withdrawal and limits after irreversible aggregation. Pseudonymous histories remain personal data. Release approved code, permitted catalogue records, synthetic fixtures and sufficiently de-identified summaries; control contacts, consent and restricted audio. FAIR reuse permits explicit access conditions[21]. Publish full participant flow, primary tests, policy comparisons, nulls and implementation disagreements. Keep existing architecture outputs, proposed choices and future observations distinct; schematic figures must not be presented as empirical panels.

This is prospective Stage 1 preparation, not completed preregistration, ethics approval or Registered Report acceptance[17, 26]. A frozen executable analysis, justified power calculation, institutional responsibility, consent, compensation and approvals are prerequisites. The companion measurement and inference papers supply separate specifications, not corroborating results. Editable text/vector sources accompany this edition; no participant records, restricted audio or deployment credentials are included. Amaç Erdem is the proposed sole author and MI's originator; final CRediT roles, affiliation, funding and interests require author verification and are not inferred from an email address. Generative-AI assistance supported editorial work, literature retrieval, figure code and production; scientific approval and journal disclosure remain the author's responsibility. This compression adds no empirical evidence.

References

[1] Vuust, P., Heggli, O. A., Friston, K. J. & Kringelbach, M. L. Music in the brain. *Nature Reviews Neuroscience* **23**, 287–305 (2022). doi:10.1038/s41583-022-00578-5.

[2] Cheung, V. K. M. et al. Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology* **29**, 4084–4092.e4 (2019). doi:10.1016/j.cub.2019.09.067.

[3] Gold, B. P. et al. Predictability and uncertainty in the pleasure of music: a reward for learning? *Journal of Neuroscience* **39**, 9397–9409 (2019). doi:10.1523/JNEUROSCI.0428-19.2019.

[4] 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). doi:10.3758/s13415-013-0161-y.

[5] Salimpoor, V. N. et al. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience* **14**, 257–262 (2011). doi:10.1038/nn.2726.

[6] Salimpoor, V. N. et al. Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* **340**, 216–219 (2013). doi:10.1126/science.1231059.

[7] Gold, B. P. et al. Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proceedings of the National Academy of Sciences USA* **116**, 3310–3315 (2019). doi:10.1073/pnas.1809855116.

[8] Gold, B. P. et al. Auditory and reward structures reflect the pleasure of musical expectancies during naturalistic listening. *Frontiers in Neuroscience* **17**, 1209398 (2023). doi:10.3389/fnins.2023.1209398.

[9] Erdem, A. Musical Intelligence: A Glass-Box Architecture for Computational Models of Music Listening. Author manuscript, 8 October 2026. Dated architecture record.

[10] Norman-Haignere, S. V. et al. Multiscale temporal integration organizes hierarchical computation in human auditory cortex. *Nature Human Behaviour* **6**, 455–469 (2022). doi:10.1038/s41562-021-01261-y.

[11] Bonetti, L. et al. Spatiotemporal brain hierarchies of auditory memory recognition and predictive coding. *Nature Communications* **15**, 4313 (2024). doi:10.1038/s41467-024-48302-4.

[12] Quiroga-Martinez, D. R. et al. Reduced prediction error responses in high- as compared to low-uncertainty musical contexts. *Cortex* **120**, 181–200 (2019). doi:10.1016/j.cortex.2019.06.010.

[13] Mencke, I., Omigie, D., Wald-Fuhrmann, M. & Brattico, E. Atonal music: can uncertainty lead to pleasure? *Frontiers in Neuroscience* **12**, 979 (2019). doi:10.3389/fnins.2018.00979.

[14] Lehne, M., Rohrmeier, M. & Koelsch, S. Tension-related activity in the orbitofrontal cortex and amygdala: an fMRI study with music. *Social Cognitive and Affective Neuroscience* **9**, 1515–1523 (2014; online 2013). doi:10.1093/scan/nst141.

[15] Noboa, M. de L., Kertész, C. & Honbolygó, F. Neural entrainment to the beat and working memory predict sensorimotor synchronization skills. *Scientific Reports* **15**, 10466 (2025). doi:10.1038/s41598-025-93948-9.

[16] Yuan, Y. et al. Decoding auditory working memory load from EEG alpha oscillations. *Psychophysiology* **62**, e70210 (2025). doi:10.1111/psyp.70210.

[17] Nature Human Behaviour. Registered Reports: submission guidelines. Publisher guidance, accessed 10 October 2026.

[18] Hawthorne, C. et al. Enabling factorized piano music modeling and generation with the MAESTRO dataset. *ICLR* (2019). Dataset, versions and licence.

[19] Woods, K. J. P., Siegel, M. H., Traer, J. & McDermott, J. H. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* **79**, 2064–2072 (2017). doi:10.3758/s13414-017-1361-2.

[20] Bridges, D., Pitiot, A., MacAskill, M. R. & Peirce, J. W. The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ* **8**, e9414 (2020). doi:10.7717/peerj.9414.

[21] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016). doi:10.1038/sdata.2016.18.

[22] Mas-Herrero, E. & Marco-Pallarés, J. Predictive processes shape individual musical preferences. *Proceedings of the National Academy of Sciences USA* **122**, e2500494122 (2025). doi:10.1073/pnas.2500494122.

[23] Gu, Y., Shao, C., Li, J. & Fan, Y. Interpretable deep learning reveals distinct spectral and temporal drivers of perceived musical emotion. *Scientific Reports* **16**, 4205 (2026). doi:10.1038/s41598-025-34238-2.

[24] Albury, A. W., Bianco, R., Johnson, A. P. & Penhune, V. B. Stimulus predictability and liking enhance auditory–motor encoding and memory for melodies. *Annals of the New York Academy of Sciences* **1560**, e70310 (2026). doi:10.1111/nyas.70310.

[25] Wilson, R. C. & Collins, A. G. E. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019). doi:10.7554/eLife.49547.

[26] Music Perception. Registered Reports: guidelines for authors. Official author guidelines (accessed 10 October 2026).