

Measuring musical experience across time: the MI-Lab framework

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

Abstract. MI-Lab specifies how online music research can preserve the identity, timing and informational context of each observation. A measurement model separates experience-related variation from reporting delay, scale use and temporal dependence. Immutable releases join recordings, personal calibration, schedules and prepared predictions; separate clocks and exposure identifiers distinguish listening from data transport. Chronological replay evaluates updates without changing collection history. Operational tests qualify acquisition, recovery and reproducibility. A future neural extension requires independent acquisition, physical synchronization and modality-specific observation models. This is a prospective measurement specification, not a deployed laboratory or validated neural account.

No participant or neural data. Operational acceptance tests are proposed, not reported as passed. All curves are schematic.

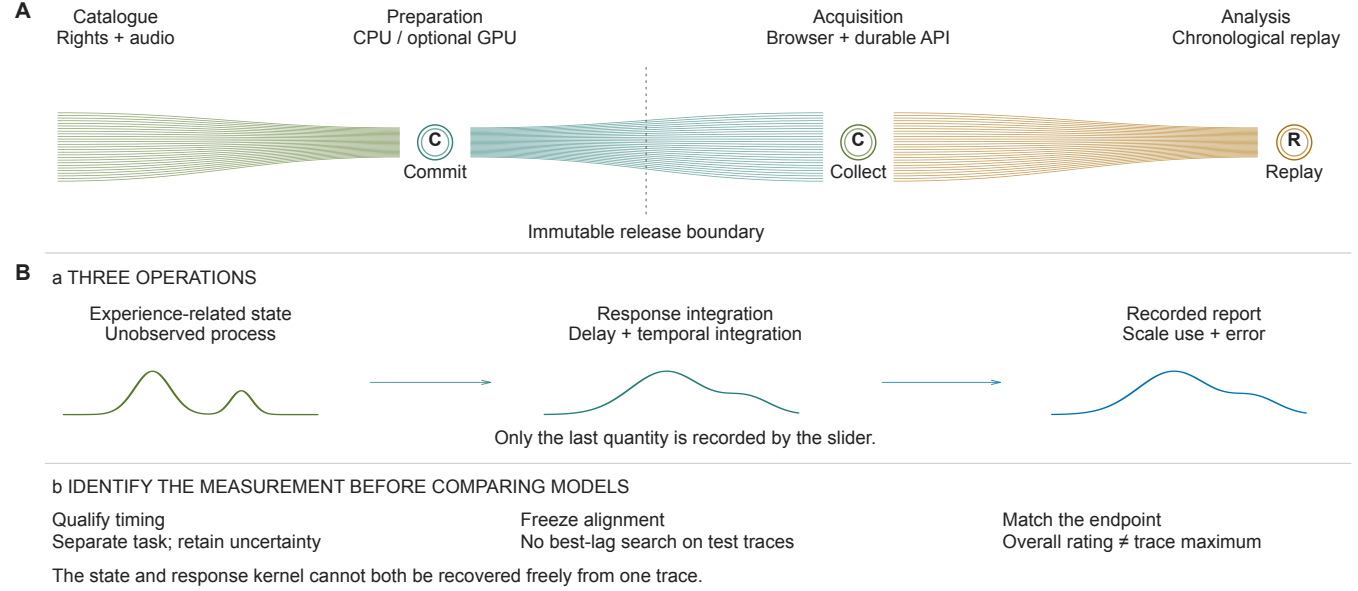


Fig. 1. Measurement architecture. **A**, Catalogue and personal preparation precede an immutable release; acquisition retrieves approved artifacts and appends observations, while analysis replays them. Curves show lineage, not traffic. **B**, Experience-related variation, response integration and recorded scale are different quantities. Independent timing qualification and fixed alignment constrain comparison; one trace cannot freely identify both a latent process and its response kernel.

Scientific target and observation model

The companion protocol asks whether personal history and expectations improve prediction of later musical pleasure[1]. Measurement must preserve what a listener heard, what they previously experienced, what a model could know and when each report was made. Listener learning, personal-model learning, policy value and predictive accuracy therefore require distinct records. Neural reward studies[5, 6, 7] motivate questions but do not turn online ratings into neural observations.

Let $a_s(t)$ be delivered audio, e an exposure and $\ell_{ise}(t)$ an unobserved experience-related process. A continuous report can be represented as

$$y_{ise}(t) = g_i \left(\int_0^\infty K_i(u) \ell_{ise}(t-u) du \right) + \epsilon_{ise}(t),$$

where K_i describes reporting delay/integration, g_i scale use and ϵ possibly autocorrelated error. This family does not equate pleasure with a single neural signal. Overall excerpt ratings may reflect duration, peaks and retrospective integration; they are not defined as trace means/maxima.

Latent experience, reporting kernel and scale transformation are not jointly identifiable from one trace without restrictions. A smooth feature may correlate with a delayed slider despite different instantaneous processes; optimizing lag on each test trace can absorb model error, while poor alignment can obscure useful information. Independently qualify report delay, retain uncertainty about its transfer to music and freeze alignment/sensitivity rules before confirmation. Calibration characterizes reporting, not direct experience. Continuous reporting changes attention/motor demands; conclusions concern listening while reporting. A randomized passive-versus-report comparison is a future experiment. Online screening/metadata cannot calibrate home headphones, rooms or hearing retrospectively.

Exposure is scientifically consequential: repeated sound can change recognition, expectation and liking[8]. Retain recording, playback, previous exposure and response phase together. Audio hashes identify content, not listener state; an upload retry preserves history, while replay creates an exposure. Dense trace samples improve recorded temporal resolution, not independent participant count.

Immutable records and acquisition semantics

Research objects. Separate catalogue preparation, personal preparation, acquisition and analysis by allowed information and immutable artifacts. A shared isolated worker decodes/verifies each permitted audio asset, computes frozen MI features and stores a versioned artifact. A personal job consumes the 20-work baseline, fits calibration, assigns the arm and prepares the complete schedule. Acquisition admits only released sessions; analysis uses pseudonymous versioned exports and explicit chronological cutoffs. The design supplies scientific comparability, not a novel database engine; standard managed services can implement it. Add orchestration clusters, vector databases or brokers only for measured needs. Provider-specific deployment, sizing, access and recovery candidates remain in the existing Supplementary Methods.

A release records protocol/consent versions, baseline digest, catalogue snapshot, engine revision, preprocessing, coefficients, seeds, selection assignments, schedule and fixed predictions. Type quantities as shared acoustic features, personal-calibration outputs, later analysis predictions or observations; a named MI surprise channel is not a surprise report. Catalogue identity includes composition, movement, arrangement, performance, master and excerpt. Hashes verify content/encoding identity; curated metadata establishes musical relations. Rights attach to exact sources and permitted derivatives. Only the paired-performance subset needs three versions per family, avoiding unnecessary restrictions on transfer coverage.

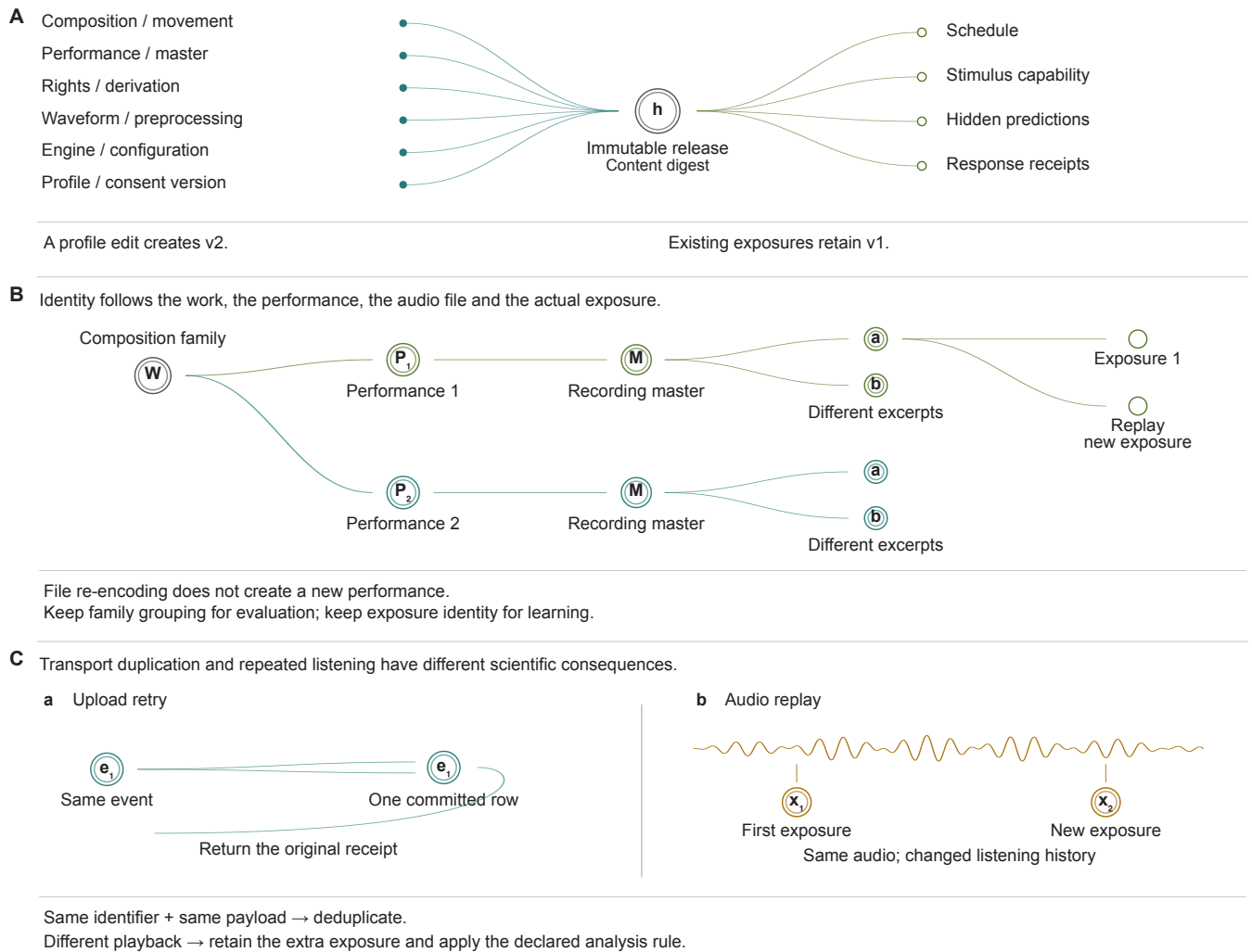


Fig. 2. Identity and provenance. **A**, A release binds audio, rights, preprocessing, engine and personal-profile versions to schedules, hidden predictions and receipts; edits create new versions. **B**, Composition, performance, recording, excerpt and exposure identities answer different questions. **C**, Identical event retries return the original receipt; playback repeats create new exposures. Digests verify identity, not anonymity, musical independence or recording rights. No acquired catalogue or deployed semantics are claimed.

Access. Stable identifiers do not imply shared credentials. The browser receives only the participant-facing manifest, short-lived audio capabilities and validated write endpoints; it receives no other profiles, hidden predictions or server keys. Analysts use research views without contacts; preparation workers get assigned baselines/audio through scoped service operations, not contact-database access. Digests do not anonymize personal data.

Instrument and archive. Pin jsPsych or equivalently audited components around Web Audio[2]. Preload/decode the next approved stimulus and record audio-context start, browser-monotonic responses, visibility changes, underruns and absolute server receipt. These clocks answer different questions; timestamp resolution is not physical timing accuracy[3]. The primary pleasure endpoint is an immediate post-excerpt rating. Sample the secondary trace at 5 Hz during 25-second transfer excerpts, batching every two seconds. Probability allocation, confidence, expected enjoyment, surprise and experienced pleasure are separate event types tied to before/after-outcome phases. Preserve intended distributions, sampled sequences, delivered outcomes and reported beliefs; scheduled probability .20 does not determine subjective $-\log_2 q(e)$.

Each immutable event has an identifier, exposure-local increasing sequence, schema version and payload hash. Keep a bounded IndexedDB queue until database commit is acknowledged. An identical identifier/payload retry returns the original receipt; a conflicting payload is rejected and audited. The browser distinguishes locally saved, uploading and acknowledged states without exposing model judgments. Connection loss need not invalidate local measurement: continue only inside authorization/storage bounds; expired credentials, queue exhaustion or missing audio pause at a safe boundary. Reload reconciles committed progress

and pending events; it never guesses whether unacknowledged sound/reporting occurred. Replayed audio creates a new linked exposure; apply first-exposure exclusions/sensitivities rather than replacing observations.

Operational contract and release lifecycle

The existing research website, manuscript records and executable MI architecture do not establish provisioned laboratory services, Supabase projects, DNS/mail, queues, consent forms or GPU workers. Interest registration is not a completed participant-management system. A versioned JSON study contract fixes the 1,000-registration cap including the proposed 60-person pilot, 20 familiar families, 1:1 U/L assignment, 24 transfer excerpts, four policies and four performance pairs; U has 24 probes and L 32. Releases refer to its digest; changes require amendments.

Registration and state. Present plain-language information before authentication. Email verification establishes account control, not consent or person uniqueness. After age/language/device screening, an atomic cap transaction with unique-account and idempotency constraints claims one slot; duplicate-person suspicions require review. Record verified, consented, baseline-in-progress/locked, preparation-queued/running, ready, active and completed states, plus ineligible, failed, expired and withdrawn branches. Only server-authorized transitions count. Consent needs version and affirmative action; locking checks distinct families, prior recognition and usable ratings. Incomplete baselines are saveable but unready. Corrections create new versions, invalidate unstarted releases and queue preparation; a baseline digest plus compare-and-swap state version rejects stale workers. After session start, record annotations/future versions without rewriting heard history.

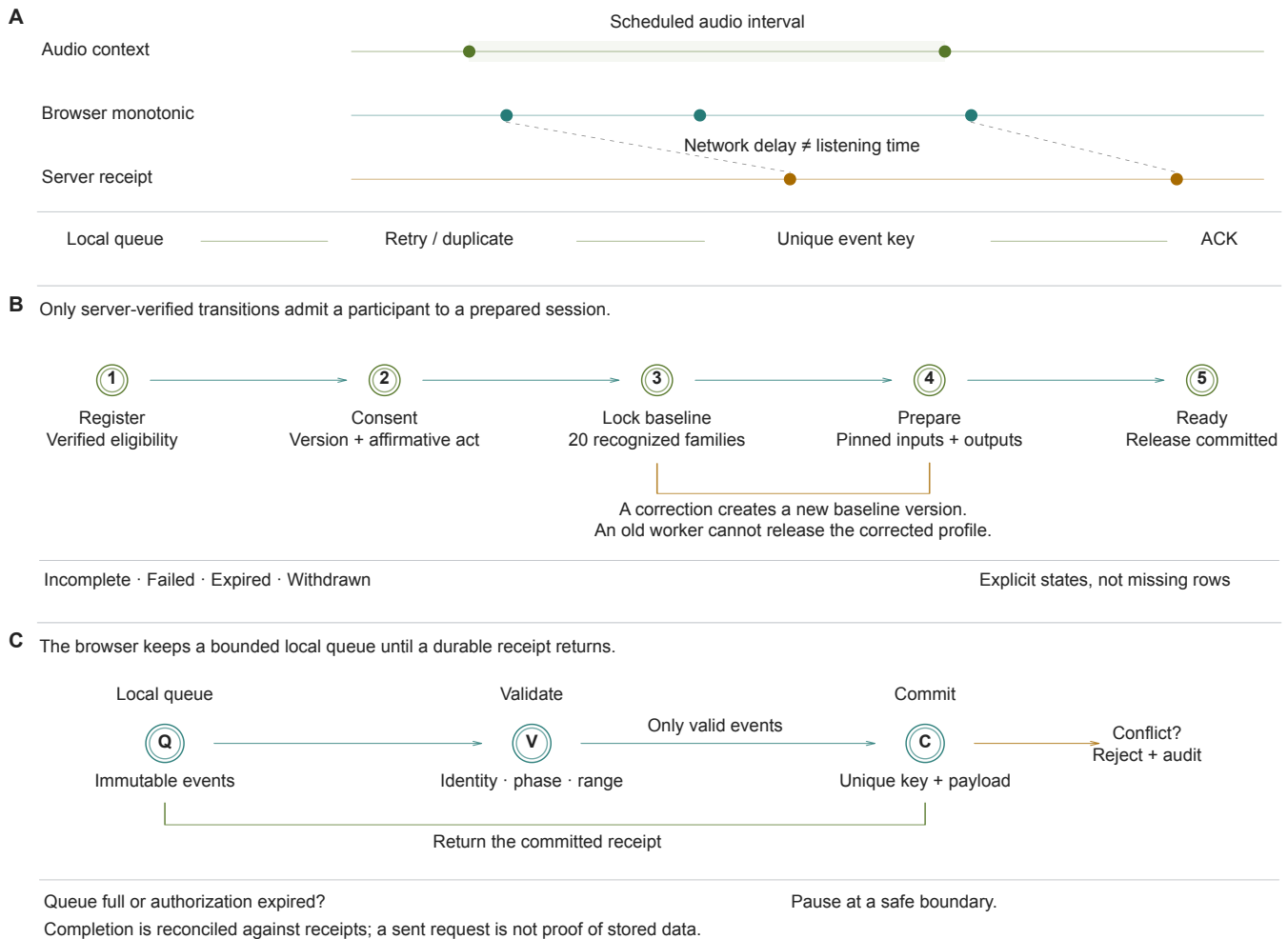


Fig. 3. Timing, readiness and durability. **A**, Audio, browser and receipt clocks remain distinct; network delay is not listening time. **B**, Registration, consent, baseline locking and preparation are separate server-authorized gates; obsolete workers cannot release corrected profiles. **C**, Durable commit precedes acknowledgement. Immutable bounded queues survive retries until receipts arrive; completion is not inferred from an optimistic interface.

Preparation and release. Baseline locking and outbox creation share a transaction. A proposed Postgres-backed queue[4] uses bounded renewable leases and expects at-least-once delivery. Deterministic job digests combine type/input versions; reuse completed artifacts only when digests and validation agree. Shared workers standardize channel layout, decoding, loudness policy and metadata checks before MI computation. Personal workers validate/fix the 20-work calibration, draw assignment/candidate partitions, rank selections and serialize hidden predictions/schedules. Optional GPU work follows explicit dependencies and cannot block live response resources. A final transaction verifies artifacts, consent/withdrawal, current baseline, catalogue permissions, complete quotas and locked model/configuration. Obsolete/withdrawn jobs may finish but cannot authorize listening. Use bounded exponential backoff with jitter, explicit retryable/terminal codes and quarantine after the limit; operators resolve failures without editing completed scientific artifacts.

Randomization and concealment. Record server-side seeds. After readiness, assign U/L equally, stratified by broad musical training; L also randomizes stable/changed distributions. Independently counterbalance outcomes, policy-slot order and performance order. Random candidate partitions share the eligible catalogue; retain assignments, inclusion rules and selection probabilities. Compute schedules before the first outcome. Expose playback order, never hidden policy scores or future predictions in payloads, analytics or filenames: disclosure would change expectation and constitute another manipulation.

Recovery and release evidence. Sample audio/browser clock mapping at session/trial boundaries, retaining offset/drift; wall-clock/timezone changes do not reorder trials. Ten 5-Hz samples form a two-second batch plus metadata. Bound local queues by bytes and unacknowledged duration; reaching either limit withholds the next trial without overwriting samples. Log visibility

loss, suspended audio, device changes and playback failure. Recovery reconstructs the last acknowledged trial and reconciles pending events. Resuming a partially heard excerpt is not a fresh exposure; protocol rules govern restart/exclusion, and all repeat playback survives export.

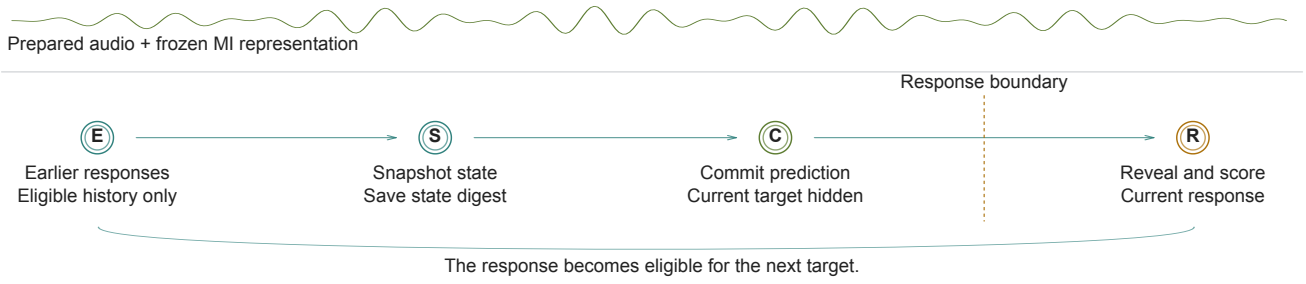
Validation, reproducibility and neural extension

Exports and replay. Feature records retain namespace, units/scaling, temporal support, engine version and semantic status. Keep acoustic coordinates, temporal summaries and cognition-labelled proxies distinguishable; raw probabilities/reports are never overwritten by transforms. Retain participant, composition, performance and exposure keys, with explicit uncollected/missing measures. A replay record joins release, ordered exposure ledger and frozen learner; before target t , expose admissible intake/earlier reports, update/snapshot state, commit prediction, then reveal/score. Save model, training cutoff, state, target and prediction digests. New learners create analysis versions, preserving collection commitments. Shared observations permit multivariate analyses and matched calibration/temporal/mechanism ablations without pretending every participant completed every assay.

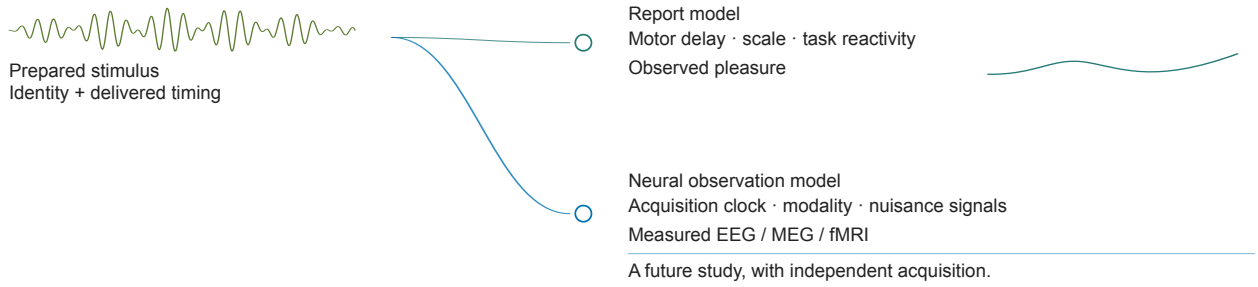
Acceptance tests. Synthetic accounts exercise registration-to-export before a human pilot. Race more than 1,000 eligible claims, including simultaneous/idempotent retries: exactly 1,000 unique successes may exist. Duplicate movements/remasters, missing recognition or too few ratings must never yield readiness. Inject duplicate jobs, lease expiry, changed baselines, removed rights and mid-preparation withdrawal: at most one exact-version release may survive, never stale authorization. Drop acknowledgements, reorder/repeat batches and conflict payloads: reconstructed events must equal fixtures with explicit conflicts and no silent duplicates.

Load-test 50 active sessions for 30 minutes, then 100 for 15, with real batch sizes/authorization. Proposed thresholds:

A Chronological replay tests a frozen learner without exposing its next answer.



B Same versioned sound · known exposure history · independently measured outcomes



Held-out neural prediction must beat matched acoustic and temporal alternatives.
A brain correlation does not identify a unique biological mechanism.

Fig. 4. Replay and independent neural measurement. **A**, Frozen stimulus features and learner specification coexist with personal-state updates from eligible earlier reports. Snapshot, commit, reveal and score; a new learner creates a new replay version, not a replacement online forecast. **B**, Behavioural and neural observation processes share stimulus identity but require different clocks, kernels, nuisance controls and independent validation. The depicted trace is not a brain recording.

95th-percentile committed acknowledgement below one second, zero lost acknowledged events and less than .1% error excluding intentionally invalid requests. Qualify physical audio/response timing separately on representative devices; network latency cannot substitute. A network trace must show zero MI/GPU/LLM inference during prepared sessions. Restore a complete synthetic subset with matching event/object hashes, including withdrawal after snapshot; reproduce earlier-only replay predictions/scores. Permission checks, delivery reconciliation and source/asset licence review complete release evidence. Report failures by affected capability, not an aggregate pass badge. None of these tests is claimed completed.

Pilot. The 60-registration pilot, included in the cap and excluded from confirmation, assesses catalogue coverage, intake effort, comprehension, timing, missingness, preparation throughput and attrition, and fixes model basis/hyperparameters. Freeze protocol amendment, catalogue, learner, environment, consent and analysis together. Dashboards show state counts/failures without hidden predictions or personal responses; stage recruitment by measured throughput/burden and disclose pauses/support. Recoverability does not require a claimed 24-hour service.

Neural observations. A separately approved EEG/MEG/fMRI study needs physical audio synchronization to acquisition clocks, qualified markers, device delay/drift and dropped-event rules. Retain recording/exposure identity while declaring new measurement resolution and assumptions. An illustrative family is $b_{iv}(t) = [\mathcal{H}_v * f_{iv}(z_s, c_{ise})](t) + \eta_{iv}(t)$, where b is measured activity, z frozen features, c admissible context and $\mathcal{H}_v$ a modality-specific operation. Prespecified fMRI haemodynamics and EEG/MEG forward/temporal assumptions differ from a slider kernel. Cortical integration evidence[9] motivates explicit support, not assignment of MI windows to regions.

Require three distinct comparisons: representation gain beyond acoustics/autocorrelation; personalization gain using the same features/observations; and pleasure-related association after timing, salience, reporting/motor and task controls. Match readout capacity, folds and tuning on independently held-out recordings/people. Choose regions, windows, preprocessing, nuisance terms and correction families without target optimization. Correlated subspaces may prevent unique attribution; estimate reliability/noise ceilings under declared assumptions. Prediction does not establish biological mechanism[10]; code interventions test computation, randomized stimuli test the assigned manipulation, and specific neural/transmitter accounts require biological discrimination. Use BIDS-compatible metadata[11], consent-sensitive access and a new sample-size/measurement rationale. This extension is not implemented neural evidence.

Status, scope and declarations. Correct provenance cannot rescue invalid constructs, and heterogeneous delivery/task-dependent reporting limit inference. Timing, recovery, privacy and reproducibility need measured results before an empirical methods claim. Institutional responsibility, ethics, compensation and consent precede recruitment; no new human data, completed validation, submission or registration is claimed. Editable sources accompany this edition; restricted audio, participant records and credentials are excluded. Amaç Erdem is the proposed sole author/MI originator; final CRediT, affiliation, funding and interests need verification, not inference from email. AI-assisted editorial/literature/figure/document work does not validate the laboratory; the author approves science and disclosure.

References

- [1] Erdem, A. Individual expectations and adaptive prediction of musical pleasure. Author manuscript, journal preparation revision 1, 10 October 2026. Distributed with this edition; not a published result.
- [2] jsPsych. Version 8 documentation: running experiments and plugin API. Experiment execution; core API (accessed 10 October 2026).
- [3] 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.
- [4] Supabase. Supabase Queues. Official documentation (accessed 10 October 2026).
- [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] Ferreri, L. et al. Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences USA* **116**, 3793–3798 (2019). doi:10.1073/pnas.1811878116.
- [8] 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.
- [9] 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.
- [10] Kriegeskorte, N. & Douglas, P. K. Cognitive computational neuroscience. *Nature Neuroscience* **21**, 1148–1160 (2018). doi:10.1038/s41593-018-0210-5.
- [11] Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* **3**, 160044 (2016). doi:10.1038/sdata.2016.44.