

Measuring musical experience across time: the MI-Lab framework

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

Abstract A continuous pleasure rating reflects musical experience, response delay and the demands of reporting. MI-Lab is a proposed framework for relating that response to the sound heard and the information available to a predictor. Its observation model makes reporting delay, scale use and temporal dependence explicit. Versioned records link recordings, personal calibration, exposure history, schedules and predictions, while separate audio, browser and server clocks preserve their timing. Chronological replay permits evaluation of personal-model updates without rewriting the listening sequence. The framework specifies acquisition, recovery and reproducibility tests that must precede empirical use. A neural extension would require independently recorded activity, physical synchronization and modality-specific observation models. This paper presents the measurement design and its validation requirements; operational and participant results are not yet available. Prospective author manuscript. All figures are conceptual or analytic illustrations.

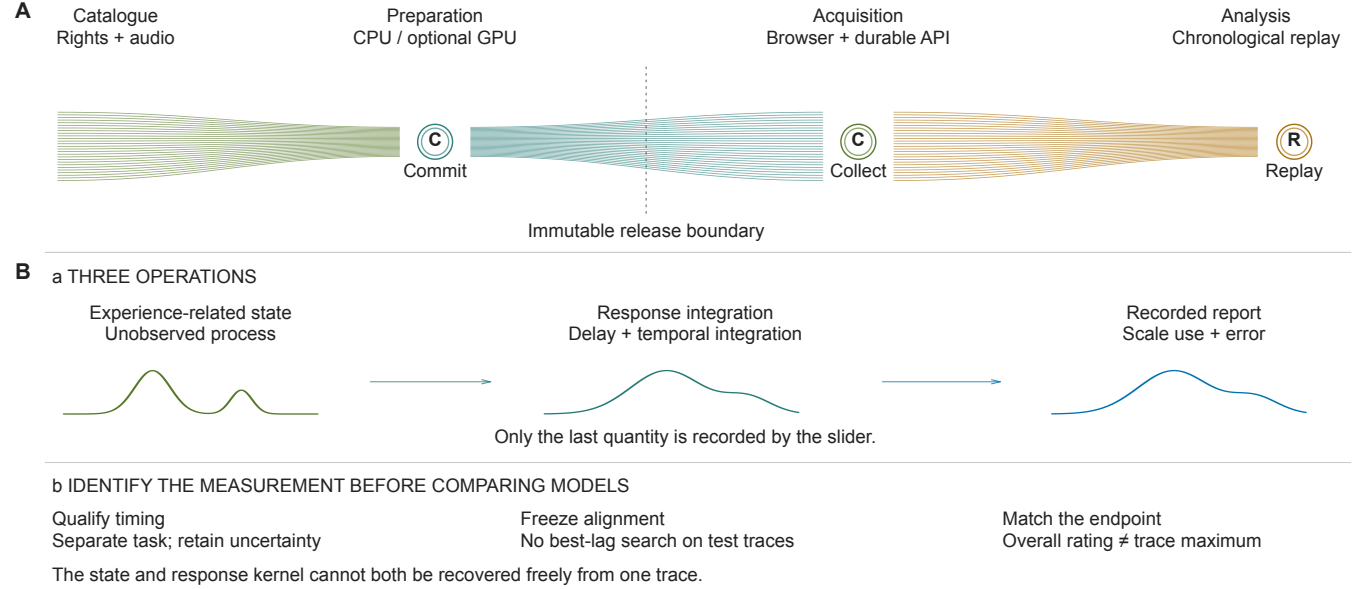


Fig. 1. Measurement framework **A**, Catalogue and personal preparation produce a fixed, versioned session. Acquisition retrieves its approved artifacts and appends observations; analysis can subsequently replay the record. Curves indicate information flow rather than measured traffic. **B**, Experience, response integration and recorded scale are distinct. Independent timing assessment and fixed alignment rules constrain comparison. A single trace cannot identify both an unrestricted latent process and its response kernel.

What a continuous response measures

A pleasure rating depends on the sound, the listener's previous experience and the act of reporting. The companion protocol asks whether personal history and expectations improve prediction of later musical pleasure[1]. Answering that question requires a record of what was heard, when it was heard, when a response was made and which information was available to the predictor. MI-Lab is a proposed framework for maintaining these relationships from stimulus preparation through analysis.

Let $a_s(t)$ denote delivered audio, e an exposure and $\ell_{ise}(t)$ an unobserved experience-related process. A continuous report can be expressed 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 represents reporting delay and integration, g_i represents scale use, and ϵ allows temporally correlated error. This formulation distinguishes the experience from its recorded response. An overall excerpt rating may also reflect duration, peaks and retrospective integration, so it is not defined as the mean or maximum of the trace.

One trace cannot identify the latent process, response kernel and scale transformation without additional restrictions. For example, a smooth audio feature may correlate with a delayed slider even when their instantaneous dynamics differ. Fitting a separate lag to each test trace could absorb prediction error; ignoring delay could obscure a useful model. Reporting delay will therefore be qualified independently, with alignment and sensitivity rules fixed before confirmation. Uncertainty about whether calibration transfers to musical listening will remain part of the analysis.

Reporting also imposes attentional and motor demands. The resulting inference concerns listening while reporting; a passive-

versus-reporting comparison would require a separate randomized experiment. Home equipment and screening metadata cannot retrospectively calibrate headphones, room acoustics or hearing. Neural reward studies[5, 6, 7] motivate related questions but require measurements beyond these reports.

Exposure history is equally important. Repetition can alter recognition, expectations and liking[8]. A recording hash identifies content, not the listener's state. Uploading the same event again leaves the exposure unchanged, whereas replaying the sound creates another exposure. The framework therefore links recording identity, playback, previous exposure and response phase. Frequent trace samples provide finer temporal sampling within that record, not additional independent participants.

Stimuli, observations and access

Preparation and versioning The design separates catalogue preparation, personal preparation, acquisition and analysis. An isolated shared worker verifies and decodes permitted recordings, computes frozen MI features and saves versioned artifacts. A personal preparation job uses the 20-work baseline to fit calibration, assign the arm and prepare the full schedule. Acquisition accepts only released sessions, while analysis operates on pseudonymous exports with explicit chronological cutoffs. Standard managed services can support this workflow; additional clusters, vector databases or message brokers would require a demonstrated operational need. Provider and deployment candidates are described in the existing Supplementary Methods.

A released session records the protocol and consent versions, baseline digest, catalogue snapshot, engine revision, preprocessing, coefficients, seeds, selection assignments, schedule and fixed predictions. Shared acoustic features, personal calibration, later analysis predictions and observations retain separate types. Thus, an

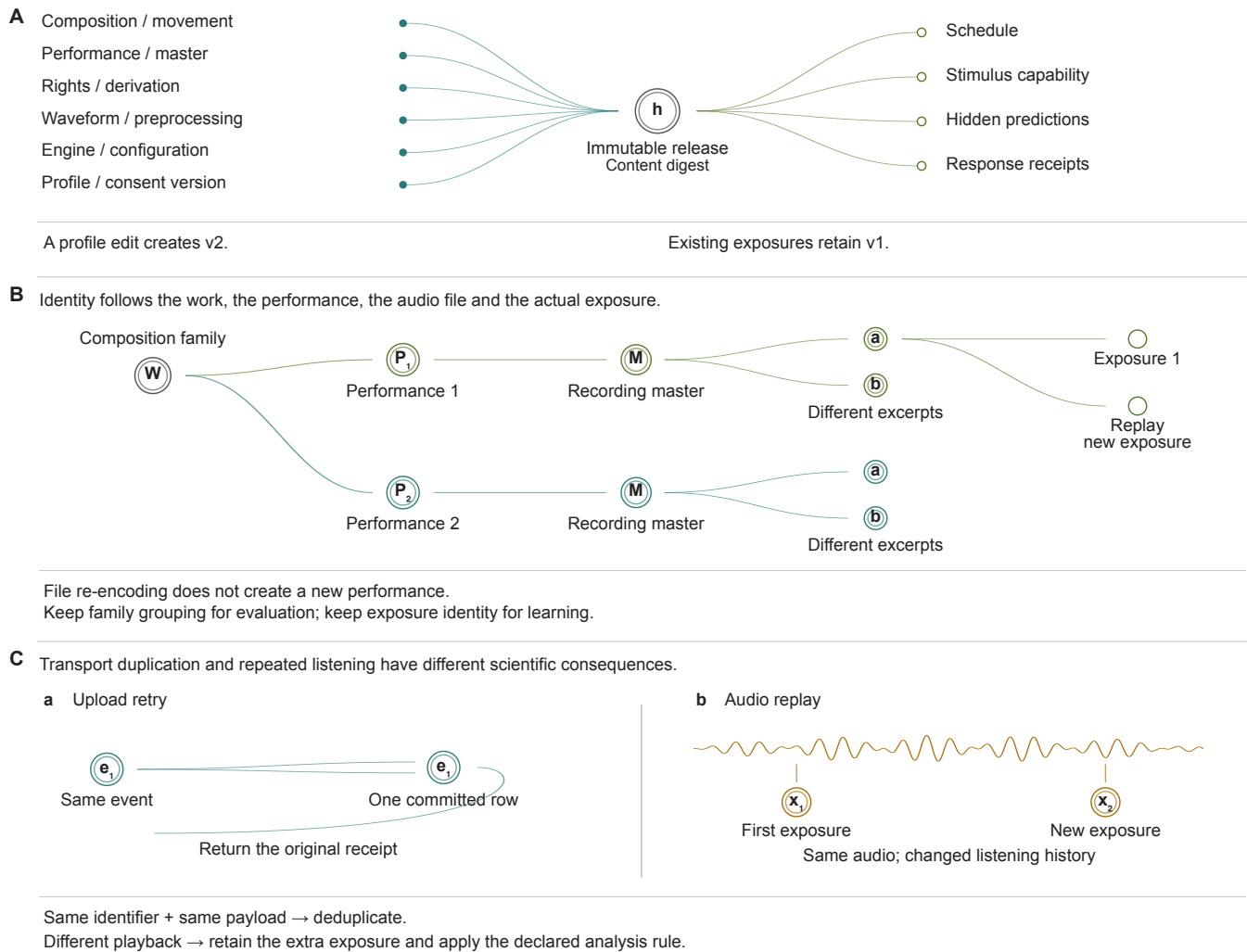


Fig. 2. Identity and provenance **A**, A release connects audio, permissions, preprocessing, engine and personal-profile versions with schedules, hidden predictions and receipts. Edits create new versions. **B**, Composition, performance, recording, excerpt and exposure identifiers distinguish musical content from its presentation. **C**, Identical event retries return the original receipt; repeated playback creates a new exposure. Digests establish identity, not anonymity, musical independence or recording rights. These are proposed records and operations.

MI channel named surprise cannot be substituted for a participant's surprise report.

Catalogue records distinguish composition, movement, arrangement, performance, master and excerpt. Content hashes establish file or encoding identity, while curated metadata establishes musical relationships. Permissions attach to exact sources and derivatives. Only the paired-performance subset requires three versions per family, leaving the wider transfer catalogue less restricted.

Access to research records The browser receives the participant-facing manifest, short-lived permissions to retrieve audio and validated response endpoints. It does not receive other profiles, hidden predictions or server credentials. Analysts work with research views without contact information. Preparation workers access only the baselines and audio assigned to them, rather than the contact database. Stable identifiers and hashes support linkage and integrity, but do not anonymize personal information.

Audio and response recording The browser instrument will use a pinned version of jsPsych, or equivalently audited components, with Web Audio[2]. It will preload and decode the next approved stimulus and record audio-context start, browser-monotonic response times, visibility changes, underruns and absolute server receipt. Each clock serves a different purpose. A precise timestamp alone does not establish physical timing accuracy[3].

The primary pleasure endpoint is an immediate post-excerpt rating. During each 25-second transfer excerpt, a secondary trace will be sampled at 5 Hz and uploaded in two-second batches. Probability allocation, confidence, expected enjoyment, surprise and experienced pleasure are separate event types, assigned to their before- or after-outcome phase. Intended distributions, sampled sequences,

delivered outcomes and reported beliefs remain distinct; a scheduled probability of .20 does not determine the listener's $-\log_2 q(e)$.

Reliable event storage Each event has an identifier, an increasing sequence number within its exposure, a schema version and a payload hash. A bounded IndexedDB queue retains events until database commit is acknowledged. An identical identifier and payload returns the original receipt when retried; a conflicting payload is rejected and logged. Participants will see whether responses are saved locally, uploading or acknowledged, without seeing model judgments.

Connection loss can be tolerated while audio, authorization and local storage remain available. Expired credentials, a full queue or missing audio will pause the session at a safe boundary. After reload, the system will reconcile committed progress with pending events. Uncertainty about unacknowledged playback or reporting will remain explicit. Replayed sound creates a linked exposure, and first-exposure exclusions or sensitivity analyses apply without replacing earlier observations.

Session preparation and recovery

The proposed study specification is a versioned JSON record containing the 1,000-registration cap, including the 60-person pilot, 20 familiar families, 1:1 U/L assignment, 24 transfer excerpts, four selection policies and four performance pairs. Arm U contains 24 probes and arm L contains 32. Releases refer to this record's digest; changing the specification requires an amendment. The research website and executable MI architecture provide starting components, while participant management, consent, queues, worker services and operational testing remain to be implemented and qualified.

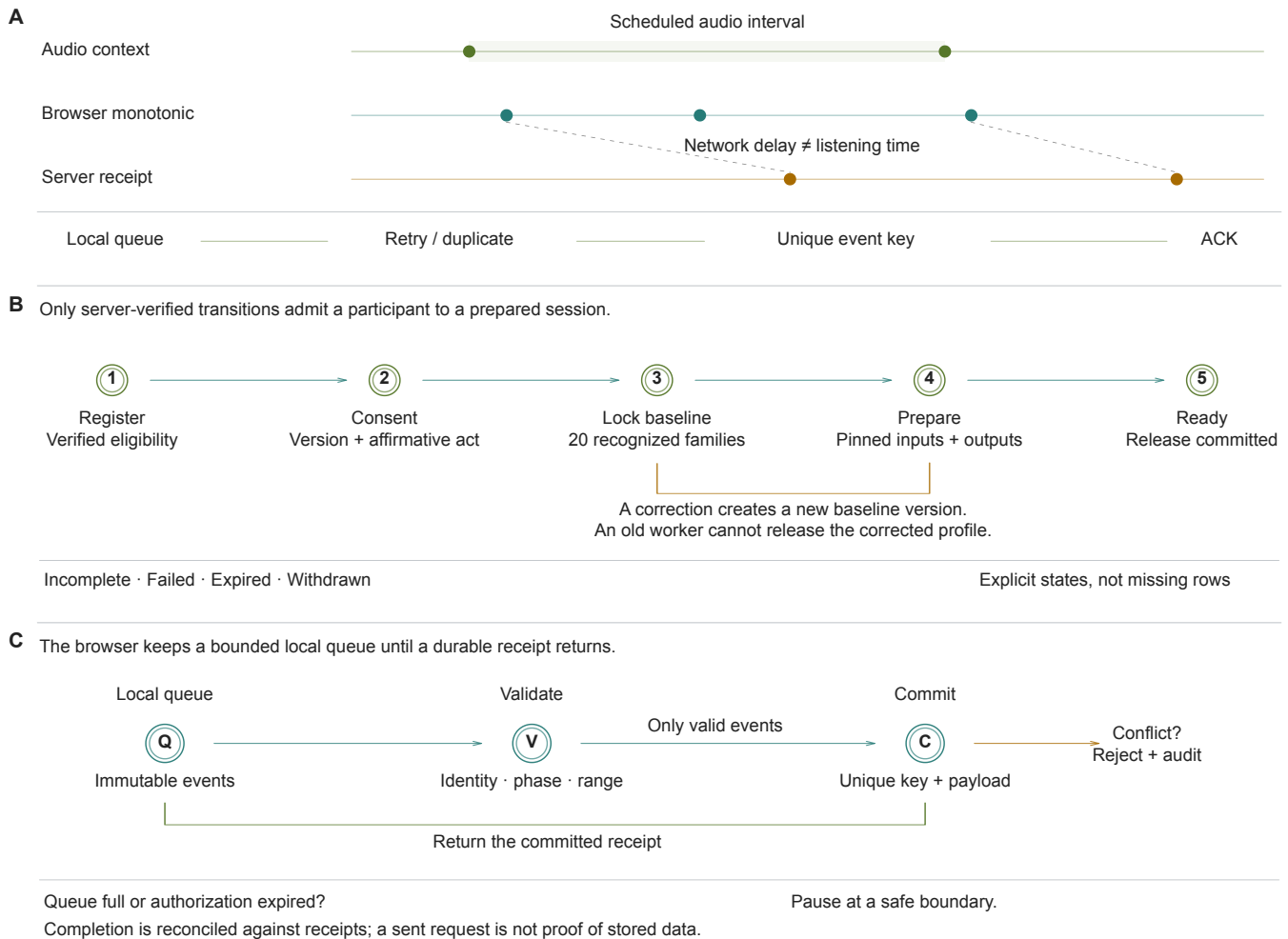


Fig. 3. Timing and recovery **A**, Audio, browser and server-receipt clocks describe different events; network delay does not measure listening time. **B**, Registration, consent, baseline locking and preparation are separate authorized transitions. A worker using an obsolete profile cannot release a corrected session. **C**, Database commit precedes acknowledgement. Bounded local queues retain events until their receipts arrive, so interface state alone does not establish completion.

Registration and consent Plain-language study information will precede authentication. Email verification establishes control of an account, rather than consent or proof of a unique person. After age, language and device screening, an atomic transaction will reserve one place under the cap. Unique-account and idempotency constraints prevent duplicate claims; suspected duplicate people require review.

The server will record verified, consented, baseline-in-progress, baseline-locked, preparation-queued, preparation-running, ready, active and completed states, together with ineligible, failed, expired and withdrawn states. Consent requires its own version and affirmative action. Baseline locking checks distinct families, prior recognition and usable ratings; incomplete baselines remain saveable but cannot become ready.

Corrections create a new baseline version, invalidate an unstarted release and queue new preparation. A baseline digest and compare-and-swap check on the state version prevent an obsolete worker from releasing it. Once listening has begun, corrections will be retained as annotations or future versions, preserving the record of what was actually heard.

Preparation jobs Locking a baseline and recording its pending preparation job will occur in one transaction. A proposed Postgres-backed queue[4] uses renewable, bounded leases and assumes that jobs may be delivered more than once. Deterministic job digests combine job type and input versions; completed artifacts can be reused only when their digests and validation match.

Shared workers will standardize channel layout, decoding, loudness policy and metadata checks before computing MI features. Personal workers will validate and fit the 20-work calibration, draw assignments and candidate partitions, rank selections, and save hidden predictions and schedules. Optional GPU jobs will follow explicit dependencies and remain separate from live response resources. A final transaction will check artifacts, consent and

withdrawal, the current baseline, catalogue permissions, complete quotas and locked model settings. Obsolete or withdrawn jobs may finish computation but cannot authorize a session. Failures will use bounded exponential backoff with jitter, explicit retryable or terminal codes, and quarantine after the retry limit. Operational repairs will preserve completed scientific artifacts.

Randomization and concealment Server-side seeds will record equal U/L assignment after readiness, stratified by broad musical training. Arm L additionally randomizes stable or changed distributions. Outcome identities, policy-slot order and performance order will be independently counterbalanced. Candidate partitions share the eligible catalogue, with assignments, inclusion rules and selection probabilities retained. Schedules will be computed before the first outcome. Participant payloads, analytics and filenames must not reveal policy scores or future predictions, since this information could alter expectations.

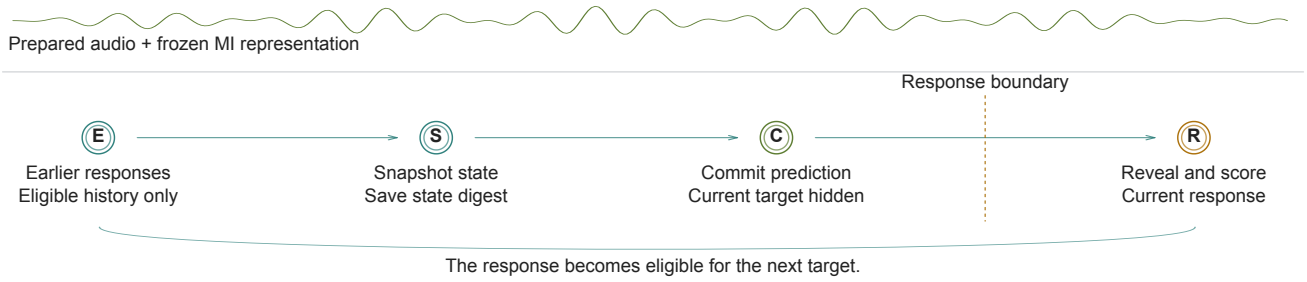
Interrupted sessions Audio-to-browser clock mapping will be sampled at session and trial boundaries to estimate offset and drift. Wall-clock or timezone changes will not reorder trials. Ten 5-Hz samples form a two-second batch plus metadata. Local queue limits will apply to both bytes and unacknowledged duration; reaching either limit withholds the next trial rather than overwriting responses.

Visibility loss, suspended audio, device changes and playback failure will be logged. Recovery reconstructs the last acknowledged trial and reconciles pending events. A partially heard excerpt remains an exposure even if restarted. Restart and exclusion rules belong to the protocol, and repeated playback remains visible in the exported record.

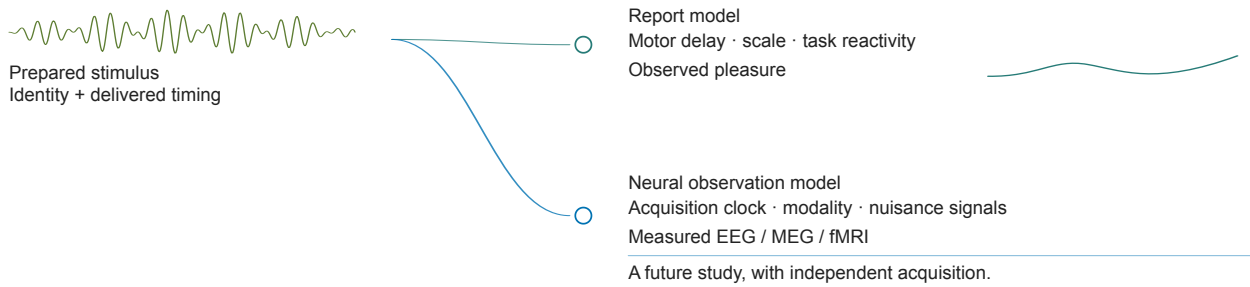
Validation and reuse

Analysis exports Each feature record will retain its namespace, units and scaling, temporal support, engine version and interpretation. Acoustic coordinates, temporal summaries and cognition-

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 neural measurement **A**, Stimulus features and the learner specification remain fixed while eligible earlier reports update personal state. The state and prediction are saved before a target is revealed and scored. A different learner creates a new analysis version. **B**, Behavioural and neural observations can share stimulus identity but need different clocks, response kernels, nuisance controls and validation. The illustrated trace is not a brain recording.

labelled proxies remain distinguishable, and derived transforms do not overwrite raw probabilities or responses. Exports include participant, composition, performance and exposure identifiers and distinguish uncollected from missing measures.

Chronological replay joins the released session, ordered exposures and a frozen learner. Before target t , the learner receives eligible intake and earlier reports; its updated state and prediction are saved before the target is revealed and scored. Records include model, training cutoff, state, target and prediction digests. A new learner creates a new analysis version rather than replacing the original collection commitments. Shared observations then support multivariate analyses and matched calibration, temporal and mechanism ablations, with the available assays explicitly recorded for each participant.

Operational tests Synthetic accounts will exercise the complete registration-to-export workflow before a human pilot. More than 1,000 eligible claims, including simultaneous and identical retries, will compete for the cap; exactly 1,000 unique successful claims should be recorded. Duplicate movements or remasters, missing recognition and insufficient usable ratings must prevent readiness. Tests will inject duplicate jobs, expired leases, revised baselines, removed permissions and withdrawal during preparation. At most one release for the exact current version may remain authorized. Dropped acknowledgements, reordered or repeated batches and conflicting payloads will test whether reconstructed events match the fixtures without silent duplication.

Load tests will run 50 active sessions for 30 minutes, followed by 100 for 15 minutes, using realistic batch sizes and authorization. Proposed thresholds are a 95th-percentile committed acknowledgement below one second, no lost acknowledged events and an error rate below .1%, excluding deliberately invalid requests. Physical audio and response timing will be qualified separately on representative devices; network performance cannot substitute for this measurement. Network traces should show no MI, GPU or LLM inference during prepared sessions.

A restore test will reconstruct a complete synthetic subset with matching event and object hashes, including a withdrawal recorded after the snapshot. Earlier-only replay predictions and scores must also reproduce. Permissions, delivery reconciliation and source licences will be checked alongside these tests. Results will be re-

ported by capability, including failures, rather than reduced to a single readiness score.

Pilot qualification The 60-registration pilot is included in the cap and excluded from confirmation. It will assess catalogue coverage, intake effort, comprehension, timing, missingness, preparation throughput and attrition, and establish the model basis and hyperparameters. Protocol amendment, catalogue, learner, environment, consent and analysis will then be frozen together. Operational dashboards will report counts and failures without displaying hidden predictions or personal responses. Recruitment will be staged according to measured throughput and burden, with pauses and support arrangements disclosed; recoverability does not imply continuous staffed availability.

Extension to neural measurements

A separately approved EEG, MEG or fMRI study would require physical synchronization between audio and acquisition clocks, qualified markers, device delay and drift estimates, and rules for dropped events. Recording and exposure identities could be reused, but the temporal resolution and observation assumptions would change. An illustrative model 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 eligible context and $\mathcal{H}_v$ a modality-specific operation. Haemodynamic assumptions for fMRI and forward or temporal assumptions for EEG/MEG differ from a slider-response kernel. Evidence for cortical integration[9] motivates explicit temporal support rather than assigning MI windows directly to brain regions.

Such a study would separate three comparisons: representation gain beyond acoustics and autocorrelation; personalization gain with features and observations held constant; and pleasure-related association after timing, salience, reporting/motor and task controls. Readout capacity, folds and tuning would be matched on held-out recordings and people. Regions, windows, preprocessing, nuisance terms and correction families would be specified without optimizing on test targets. Reliability and noise ceilings would be estimated under stated assumptions, while correlated representations may still prevent unique attribution.

Successful neural prediction would not establish a biological mechanism[10]. Code interventions test computations, random-

ized stimuli test assigned manipulations, and specific neural or transmitter accounts require evidence that distinguishes those biological alternatives. A neural extension needs BIDS-compatible metadata[11], consent-sensitive access and its own sample-size and measurement rationale.

Scope and study status

MI-Lab is a prospective measurement specification. Timing, recovery, privacy and reproducibility tests have not been reported as completed. Equipment variability and task-dependent reporting remain limitations even if the records are technically correct. Institutional responsibility, ethics review, compensation and consent must precede recruitment; interest registration alone does not constitute enrolment or an implemented laboratory.

Editable sources accompany this edition; participant records, restricted audio and credentials are excluded. Amaç Erdem is MI's originator and the proposed sole author. Affiliation, funding, competing interests and contributor statements remain to be confirmed before submission. Generative-AI assistance supported editorial work, literature retrieval, figure programming and document production. The author remains responsible for scientific approval and disclosure; this assistance provides no validation evidence for the proposed system.

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.