

Measuring musical experience across time: the MI-Lab framework

Amaç Erdem

Correspondence: amace@bu.edu · ORCID 0009-0000-1711-0036

Abstract

Temporal and personalized models of musical experience require measurements whose identity, timing and informational context remain recoverable. This article develops MI-Lab as a methodological framework for online music research, with an explicitly separate route to future neural assays. An observation model distinguishes experience-related variation from response delay, scale use and temporal dependence. A versioned release links composition, performance, recording, exposure, participant calibration and prepared predictions. Separate audio, browser and receipt clocks prevent transport timing from being interpreted as listening time. Event identity distinguishes repeated delivery from repeated listening, while chronological replay permits personal-model updates without changing the original stimulus sequence. The framework specifies measurement qualification, durable acquisition, failure recovery and matched comparison of computational accounts. A prospective neural extension requires physical stimulus synchronization, a modality-specific observation model and independently held-out EEG, MEG or fMRI evidence; online pleasure ratings do not supply these measurements. Provider-specific implementation choices are confined to supplementary methods. The contribution is a testable measurement and provenance specification, rather than a claim of a deployed laboratory or validated neural mechanism. Its utility must be established by timing, recovery, reproducibility and empirical model-discrimination studies.

Prospective methods framework · No neural or participant data · Operational acceptance tests are specified, not reported as passed.

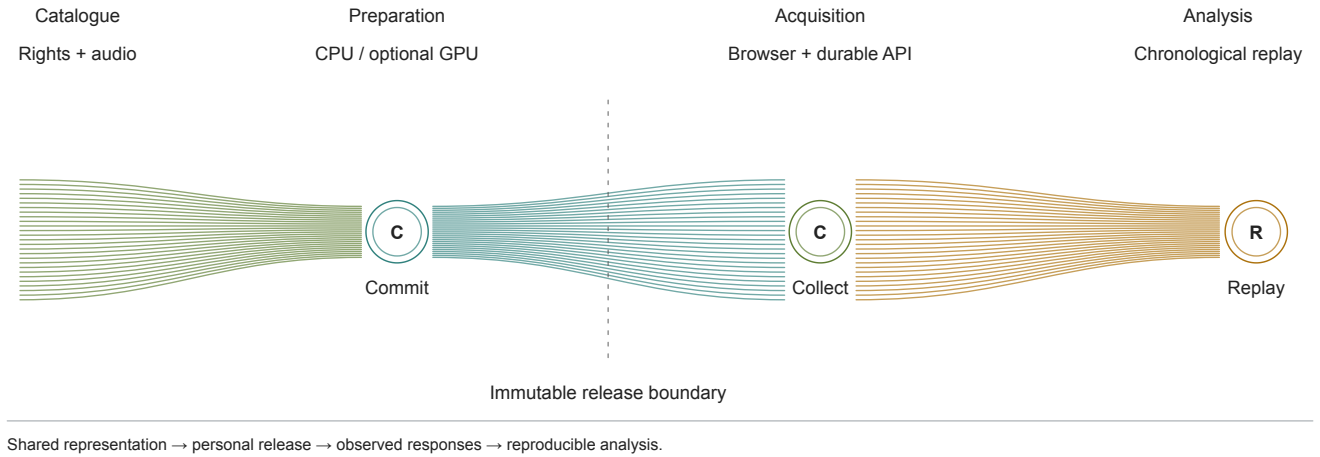


Fig. 1 | A release boundary separates preparation from observation. Shared catalogue processing and personal preparation run before a participant is admitted to listening. Acquisition then retrieves approved artifacts and appends responses. Analysis can update models through chronological replay, while retaining the original schedule and predictions. Curves depict artifact lineage rather than measured traffic. Heavy computation is absent from the listening path.

Introduction

The central scientific question is whether a personal model learned from selected music helps explain and predict subsequent pleasure. Answering it requires linking acoustic structure, temporal context, expectation, uncertainty and responses across many listeners. A web form can collect a rating; a scientific platform must also establish which information was available before that rating. MI-Lab treats this temporal boundary as a first-class data object.

The companion protocol[1] separates listener learning from model learning and policy value from prediction accuracy. The infrastructure preserves those distinctions in storage and execution. An updated model cannot overwrite a committed prediction, a repeated upload cannot become a second observation, and an edited personal profile cannot silently change a session already released. These properties determine whether a later statistical comparison is interpretable.

Measurement model and scientific target

A report is an observation of an experience

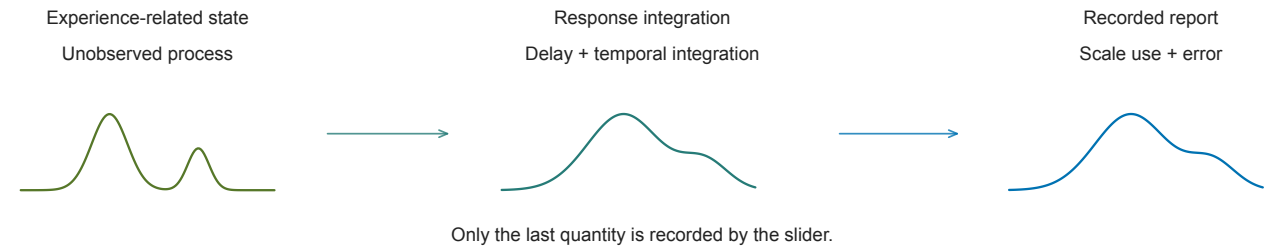
Music neuroscience relates a structured acoustic stimulus to expectation, affect, memory and neural activity. These quantities occupy different observational levels. Studies of anticipation and peak musical emotion[5], auditory-reward interactions[6] and pharmacological modulation of musical reward[7] used evidence beyond a preference rating. MI-Lab's online application measures reports and behaviour. Its contribution is to preserve the conditions under which those observations can later be compared with computational and neural accounts, rather than to infer those accounts from the names of model channels.

Let $a_s(t)$ denote the delivered audio for stimulus s , e the exposure, and $\ell_{ise}(t)$ an unobserved experience-related process. A continuous rating can be represented schematically as

$$y_{ise}(t) = g_i \left(\int_0^\infty K_i(u) \ell_{ise}(t-u) du \right) + \epsilon_{ise}(t), \quad (1)$$

where K_i describes response delay and temporal integration, g_i describes scale use, and the error may be autocor-

a THREE OPERATIONS



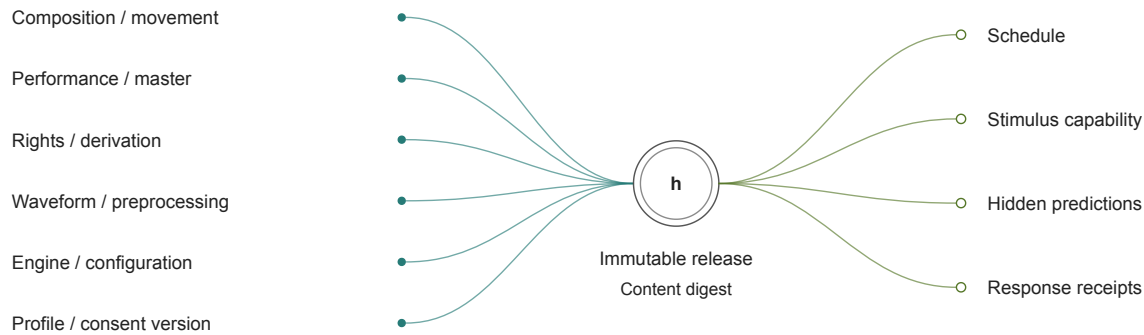
b IDENTIFY THE MEASUREMENT BEFORE COMPARING MODELS

Qualify timing	Freeze alignment	Match the endpoint
Separate task; retain uncertainty	No best-lag search on test traces	Overall rating $\neq$ trace maximum

The state and response kernel cannot both be recovered freely from one trace.

Schematic shapes only · No observed experience, rating or fitted response kernel is shown.

Fig. 2 | A continuous response requires an observation model. The unobserved experience-related process, response integration and reported scale are distinct. Shapes are schematic, without fitted parameters or participant data. Independent timing qualification and frozen alignment rules constrain inference; one trace cannot freely identify both a latent trajectory and its response kernel.



A profile edit creates v2.

Existing exposures retain v1.

A digest verifies identity · It does not anonymize personal data.

Fig. 3 | Provenance joins musical identity to computational and personal context. a, A release digest depends on audio identity, rights, preprocessing, engine configuration and a versioned personal baseline. Hidden predictions and participant-facing schedules derive from that release. Response receipts refer back to it. The digest verifies content identity; it does not anonymize personal data. b, A profile edit creates a new version. Existing exposures retain their original lineage.

related. This is an observation-model family, not a claim that subjective pleasure is a one-dimensional neural signal. An excerpt-level judgment is another measurement, potentially influenced by duration, salient moments and retrospective integration; it is not defined to equal the trace's mean or maximum.

The distinction matters for model comparison. A delayed slider can correlate with a smooth acoustic feature even when their instantaneous processes differ. Optimizing delay separately for every test trace can absorb model error. Conversely, an inaccurate delay rule can make a useful feature appear uninformative. Timing qualification, fixed alignment choices and comparison with simple temporal baselines are therefore part of the scientific method, not only interface engineering. A dense trace increases temporal resolution of recorded responses; it does not increase the number of independent listeners.

Identifiability and task-induced effects

The latent experience, response kernel and rating transformation are not jointly identifiable from one trace

without additional restrictions. In particular, temporal smoothing can be attributed either to the candidate process or to the response. MI-Lab will estimate report delay using a separate, qualified calibration task and retain uncertainty about its transfer to music. The study will use a prespecified family of alignment sensitivities rather than choose the best-fitting lag on the confirmatory outcome. Calibration establishes aspects of a reporting process, not direct access to experience.

Continuous reporting can also change attention, motor activity and listening. In the current protocol it is a common requirement, so resulting conclusions concern listening while reporting. A future randomized passive-versus-continuous-report comparison could test reactivity, but that experiment is not part of the present data claim. Online observation likewise does not normalize headphone frequency response, room conditions or hearing. Screening and metadata characterize these sources of variation; they cannot retrospectively turn home listening into calibrated acoustic delivery.

An exposure is part of the intervention

A waveform hash identifies audio content, not the listener's state. Hearing a passage again can change recognition, expectation and liking. Contemporary findings relating predictability and pleasure to motor learning and subsequent recognition illustrate why these outcomes need separate assays[8]. The recording, playback event, previous exposures and reporting phase must remain linked. A transport retry should leave that history unchanged; a replay must add an exposure. This distinction is essential whenever learning is an outcome or a predictor.

Materials and method: an immutable measurement record

The implementation has four domains: catalogue preparation, participant-specific preparation, acquisition and analysis. Their boundaries are defined by permitted information and immutable artifacts, rather than by a particular hosting provider. Supplementary Methods retain the proposed deployment, sizing, access-control and recovery details. These are operational candidates, not evidence of scientific validity.

The preparation domain uses isolated container workers. A shared job decodes and verifies each admitted audio asset, computes the frozen MI representation and stores the resulting feature artifact. A personal job consumes the versioned 20-work profile, initializes predictions, assigns the experimental arm and generates the complete schedule. The acquisition domain accepts only released sessions. The analysis domain exports pseudonymous, versioned observations and replays candidate learners under explicit information cutoffs.

These domains share stable identifiers but not unrestricted credentials. The browser receives the minimum session manifest required to run the task, short-lived access to its audio and write access through validated endpoints. It never receives other profiles, server keys or hidden model predictions. Analysts receive a de-identified research view without contact details. A preparation worker can read its assigned baseline and audio artifacts through scoped service operations; it has no reason to access the contact database.

What a personal release contains

A release is an immutable agreement between stimulus preparation and later observation. It records the protocol version, consent version, baseline digest, catalogue snapshot, engine revision, preprocessing recipe, model coefficients, seeds, selection policy assignments, schedule and fixed predictions. It also records whether a quantity is a shared acoustic feature, a personal calibration output, a future analysis prediction or an observed response. This typing prevents a model-labelled "surprise" channel from being confused with a listener's surprise report.

The catalogue distinguishes composition, movement, arrangement, performance, recording master and excerpt. Content hashes detect duplicate encodings and accidental changes; musical identity still requires curated metadata. A rights record belongs to the exact source and permitted derivative use. The corpus may contain one eligible reference for many families while only a subset supports three matched performances. Requiring three versions for every transfer item would unnecessarily constrain the broader scientific design.

MI-Lab's contribution is therefore not a novel database engine. It is the specification of how personal learning, multidimensional audio and temporally ordered observations remain comparable. Standard managed services can implement that specification if their boundaries are tested. The design intentionally avoids adding an orchestration cluster, vector database or message broker without a measured need.

The browser is an instrument, not the archive

The session runner uses a pinned version of jsPsych or equivalently audited task components around Web Audio[2]. It loads the next approved stimulus before playback and records the audio-context start, browser monotonic timestamps, visibility changes, underruns and user responses. An absolute server timestamp accompanies receipt. These clocks remain distinct. A timestamp's resolution does not establish the accuracy of the physical sound or motor response; device-level timing is qualified in the pilot[3].

The primary pleasure endpoint is an immediate post-excerpt rating. Continuous pleasure is sampled at 5 Hz during 25-second transfer excerpts and batched every two seconds. Keeping the discrete endpoint primary reduces dependence on unvalidated response lag, while the trace supports secondary tests at predeclared MI moments. Probability allocations, confidence, expected enjoyment, surprise and experienced pleasure are different event types, attached to explicit before- or after-outcome phases.

The browser maintains a bounded IndexedDB queue until the server acknowledges a committed batch. Each event has a generated identifier, monotonically increasing sequence within an exposure, a schema version and a payload hash. Re-sending an existing identifier with an identical payload returns the original receipt; a changed payload is a conflict. Acknowledgement occurs only after durable database commit. The user can see a recoverable upload state without being shown computational claims about their listening.

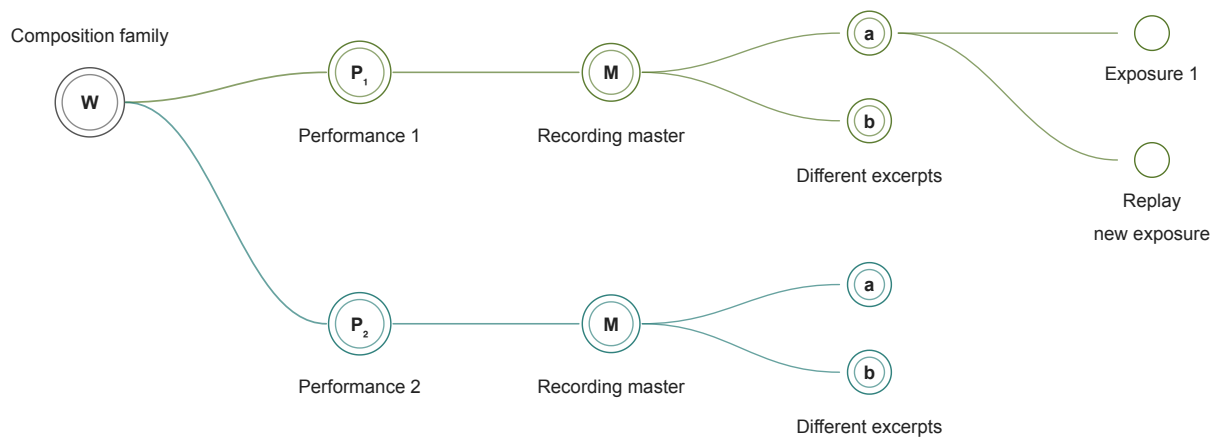
Preserve failure without manufacturing observations

A disconnected network does not necessarily invalidate local measurement. Playback and reports can continue only within the bounded, already authorized session and local storage budget. Expired authorization, exhausted queue capacity or missing audio pauses the session at a safe boundary. A reload reconstructs acknowledged progress; it does not guess whether an unacknowledged response was heard or delivered. The trial ledger resolves that ambiguity.

If an excerpt is replayed after interruption, the new exposure receives a new identifier linked to the failed attempt. The first-exposure analysis can then exclude or qualify it according to the scientific protocol. Upload duplication is a transport property; listening duplication changes the participant's experience. Collapsing both into a single "retry" would corrupt learning analyses.

The same principle applies to probabilistic tasks. The intended distribution, sampled sequence, delivered audio and reported belief survive in separate fields. Sampling an outcome with probability 0.20 does not fix its subjective surprise at $-\log_2(0.20)$: that quantity depends on the participant's prior report. This distinction makes the dataset

Identity follows the work, the performance, the audio file and the actual exposure.

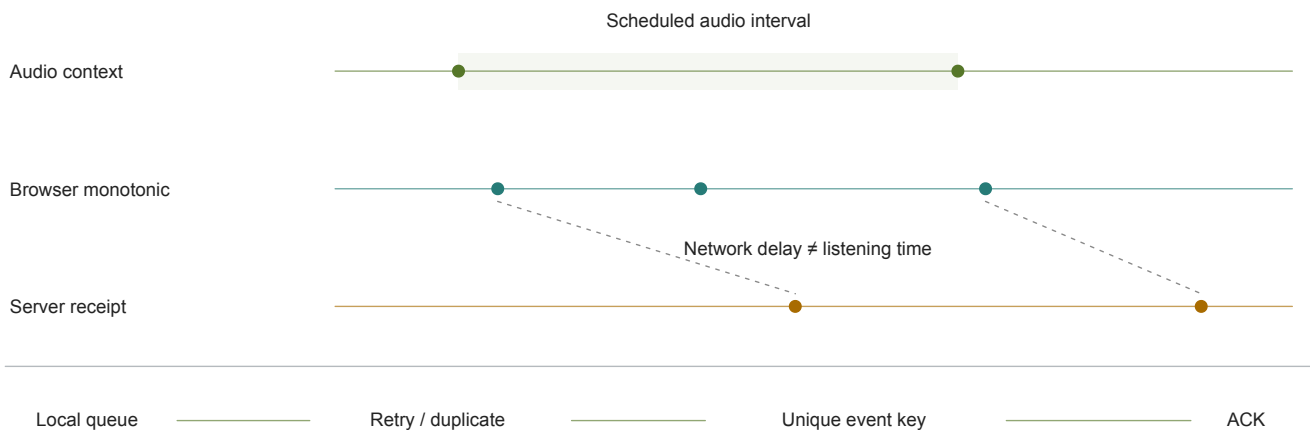


File re-encoding does not create a new performance.

Keep family grouping for evaluation; keep exposure identity for learning.

A recording digest establishes file identity, not musical independence or permission to use the recording.

Fig. 4 | Musical identity and exposure identity are different. One composition family can have multiple performances, each with its recording master and excerpts. Re-encoding a master does not create an independent performance. Hearing an identical excerpt again creates a new exposure and can change learning history. Movement and arrangement identities remain in the full catalogue record. The schematic does not describe an acquired catalogue or imply recording rights.



Repeated upload → same event · Repeated listening → a new exposure.

Fig. 5 | Reliable measurement requires separate clocks and explicit delivery semantics. a, Scheduled audio time, browser monotonic time and server receipt time answer different questions. Network delay cannot be used as a proxy for listening time. Marker locations are illustrative. b, Delivery can occur more than once; database uniqueness and payload verification preserve one logical event. Listening to the same excerpt again creates a new exposure even if its audio hash is unchanged.

useful for competing theories, without requiring them to adopt MI's assumptions.

Measurement acquisition and analysis specification

Current implementation boundary

The existing project supplies a research website, manuscript records and an executable MI architecture. This manuscript proposes the laboratory service around them. It does not establish that Supabase projects, DNS records, mail domains, production queues, consent forms or GPU workers have been provisioned. The existing interest-registration affordance must not be described as a completed participant-management system. Launch requires the operational evidence defined below.

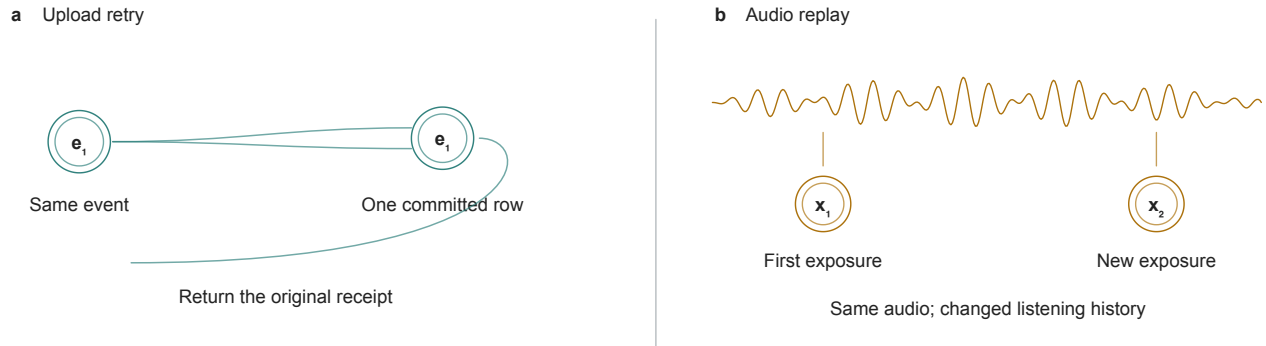
The study contract accompanies the manuscript as a versioned JSON document. It fixes the 1,000 registration

cap, proposed 60-registration pilot, 20 distinct familiar composition families, 1:1 mechanistic arm assignment, 24 transfer excerpts, four selection policies and four paired-performance trials. Arm U has 24 probe outcomes; Arm L has 32. A release always refers to one contract digest. Changing these values creates an amendment rather than silently changing an ongoing experiment.

Registration and participant state

An unauthenticated visitor first receives plain-language study information. Email verification establishes account control, not consent or scientific eligibility. After a minimal age/language/device screen, a verified eligible account claims one registration slot through a transaction with an atomic cap check. A unique account constraint and an idempotent request key prevent retries from consuming additional slots. Operational identity checks may flag suspected duplicate people for review, but the sys-

Transport duplication and repeated listening have different scientific consequences.



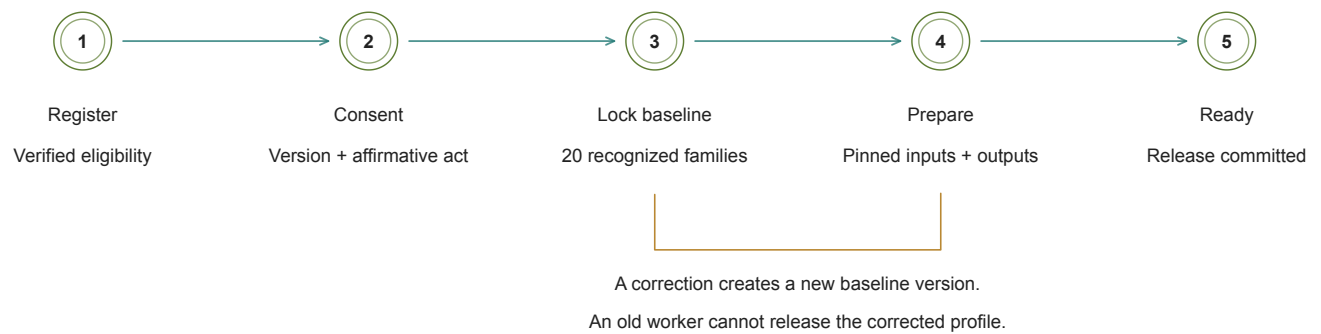
Same identifier + same payload → deduplicate.

Different playback → retain the extra exposure and apply the declared analysis rule.

A replayed response never silently replaces the first. Conflicting event payloads are rejected and audited.

Fig. 6 | A transport retry preserves an event; an audio replay creates an exposure. An identical event identifier and payload return the original committed receipt. A conflicting payload is rejected and audited. Replaying sound receives a new exposure identifier linked to the earlier attempt; the analysis handles repeated exposure under the protocol instead of silently substituting the new response. These are proposed semantics, not evidence from a deployed service.

Only server-verified transitions admit a participant to a prepared session.



Incomplete · Failed · Expired · Withdrawn

Explicit states, not missing rows

1,000 registration cap includes the 60-person pilot; it does not count completed or concurrent sessions.

Fig. 7 | Readiness is an authorized, versioned state. Verified registration, consent, baseline completion and preparation are separate gates. A locked baseline correction creates a new version, invalidates an unstarted release and queues preparation; a stale worker cannot publish readiness for that new version. Incomplete, failed, expired and withdrawn states remain explicit. The 1,000-registration cap includes the proposed 60-registration pilot and is not a count of completed datasets.

tem does not claim that email uniqueness proves person uniqueness.

The state sequence is verified registration, consented, baseline in progress, baseline locked, preparation queued, preparation running, ready, session active and completed. Ineligible, preparation failed, expired and withdrawn are explicit branches. Only server-authorized operations change state. Consent requires a version identifier and affirmative action; baseline locking validates distinct composition families, prior-recognition responses and usable calibration ratings. A person may save an incomplete baseline without being counted as ready.

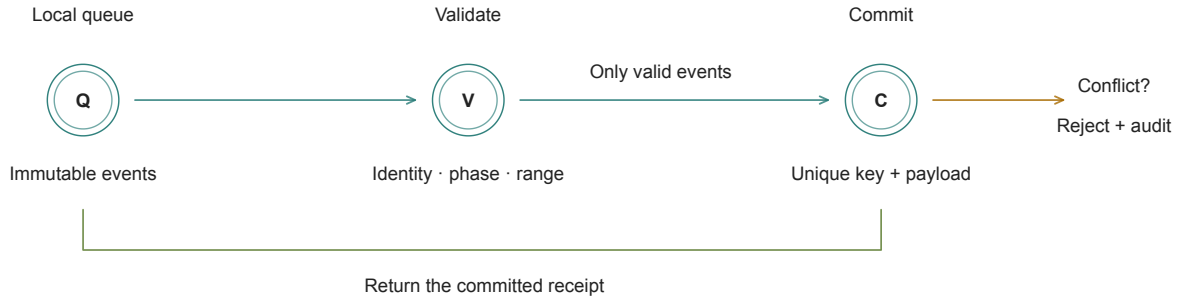
Once locked, the baseline is immutable. A permitted correction creates a new baseline version, invalidates any unstarted release and queues new personal preparation. A worker finishing the previous version cannot mark the

new participant state ready. This check uses both the baseline digest and a compare-and-swap state version. After a session starts, corrections are recorded as annotations or future versions; they do not rewrite the history that generated already heard stimuli.

Preparation queue and release transaction

The database transaction that locks a baseline also creates an outbox/queue entry, eliminating the gap between “saved” and “scheduled”. A Postgres-backed queue is the proposed durable mechanism[4]. Workers acquire bounded leases and renew them while processing. At-least-once delivery is expected, so the job identity is a deterministic digest of job type and input versions. A successfully completed artifact is reused when its digest and validation status match.

The browser keeps a bounded local queue until a durable receipt returns.



Queue full or authorization expired?

Pause at a safe boundary.

Completion is reconciled against receipts; a sent request is not proof of stored data.

Same valid event → same receipt. Changed payload → explicit conflict. No silent loss or duplicate observation.

Fig. 8 | Commit acknowledgement defines durable event acceptance. The browser retains immutable events within a bounded queue. The endpoint checks identity, ownership, schema, measurement phase, sequence and range, then attempts an atomic unique insertion with payload verification. A receipt follows durable commit, while conflicting duplicates are rejected. Exhausted queue or authorization limits pause the next trial. Completion reconciles receipts and preserves partial data rather than assuming that sending a request establishes storage.

Shared extraction and personal preparation are separate job classes. Shared extraction performs decode checks, preprocessing and MI computation. Personal preparation validates the 20-work baseline, fits the constrained personal readout, draws the assignment and candidate partition, ranks policy selections and serializes the schedule. All hidden predictions are saved before release. Optional GPU jobs run only when a dependency explicitly requires them; they never block the live response endpoint's resource pool.

Release is one transactional decision after all required artifacts exist and their digests verify. It rechecks consent, withdrawal, current baseline, catalogue permissions, complete policy quotas and the locked model/configuration version. An obsolete or withdrawn job may finish computation but cannot authorize a session. Failures use bounded exponential backoff with jitter, explicit retryable versus terminal codes and a quarantine queue after the limit. A study operator resolves terminal failures without editing completed scientific artifacts.

Randomization and hidden predictions

Assignment seeds are generated server-side and recorded with the release. The scientific design uses 1:1 assignment to U/L after readiness, stratified by broad musical training. Arm L additionally randomizes changed versus stable distributions. Outcome identities, policy slot order and paired-performance order are counterbalanced independently. All schedules are computed before the first experimental outcome.

Candidate pools for the four selection policies are randomized partitions of the same eligible catalogue for each participant. Their assignments, inclusion rules and selection probabilities are saved. The resulting policy labels and scores remain hidden from the participant; only the playback order is exposed. The API returns no future pleasure prediction in a browser payload, analytics event or downloadable filename. Revealing the model's preference could itself alter expectation and would create another experimental manipulation.

Chronological learning and multidimensional exports

Every shared feature vector has a namespace, units or scaling, temporal support, engine version and semantic status. Acoustic coordinates, temporal summaries and cognitive-labelled proxies remain distinguishable. Raw probability reports and behavioural outcomes are never overwritten by transformed values. This allows alternative definitions of uncertainty or surprise to be computed from the same observations while preserving the declared primary analysis.

The replay engine accepts a release, an ordered exposure ledger and a frozen learner specification. Before trial t , it exposes only admissible intake and earlier responses, updates the state, writes a prediction commitment, and then reveals the current target to score it. A replay record includes the model digest, training cutoff, state digest, target exposure and prediction. Replaying with a new learner creates another analysis version. It cannot replace the fixed predictions made before collection.

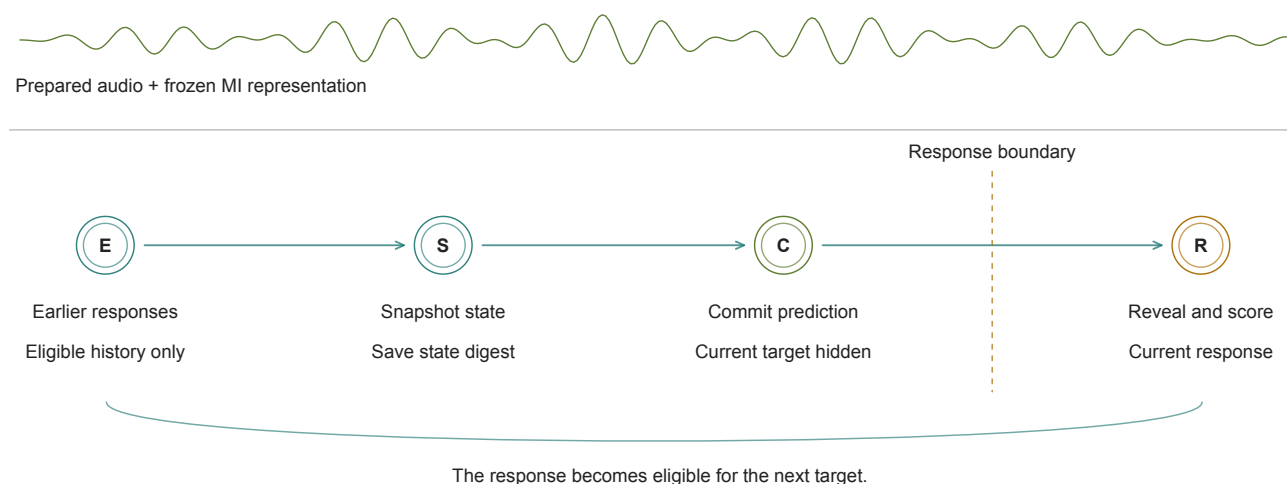
Exports retain participant, work, performance and exposure grouping keys so that dense traces are not mistaken for independent samples. A common response table supports simultaneous modelling of expectation, surprise, confidence and pleasure; explicit missingness identifies measures not collected in a task. Multivariate analysis therefore draws strength from shared provenance and overlapping observations, rather than pretending that every listener completed every possible assay. Ablations can remove personal calibration, temporal features or mechanism outputs while scoring the same held-out responses.

Timing, buffering and session recovery

Audio is preloaded and decoded before a trial begins. Playback uses a recorded Web Audio start time; interface responses use the browser's monotonic clock. Their mapping is sampled at session and trial boundaries, with offset and drift retained. Server receipt is used for delivery audit, not for reconstructing millisecond event onset. Wall-clock changes and timezone do not alter trial order.

The local queue stores immutable event records and acknowledged batch receipts. The planned 5-Hz trace pro-

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



A new learner creates a new replay version; it cannot replace predictions committed before collection.

Fig. 9 | Frozen stimulus descriptions and changing personal states coexist. The prepared audio representation and learner specification remain fixed during a replay run. Only admissible intake and earlier responses inform the next state. The state and prediction are recorded before the current outcome is revealed and scored. A new learner creates a distinct analysis version; it cannot replace fixed predictions committed before collection. This is a replay estimate, not evidence that a forecast was issued online.

duces ten samples per two-second batch, plus metadata. A queue limit is defined in both bytes and elapsed unacknowledged time. On reaching it, the next trial is withheld rather than overwriting older samples. The system reports when data are saved locally, uploading or acknowledged without promising completion before server commit.

Visibility loss, suspended audio, device changes and playback failures generate explicit quality events. A recovery controller reconstructs the last acknowledged trial and reconciles local pending events. Resuming within a partially heard excerpt is not equivalent to a fresh exposure; the scientific protocol determines whether to restart and exclude it from the first-exposure endpoint. Repeated playback remains observable in all exports.

Validation plan and anticipated failure modes

Before human pilot use, synthetic accounts exercise the entire sequence from registration to export. A race test attempts more than 1,000 eligible slot claims, with repeated idempotency keys and simultaneous requests; exactly 1,000 unique successful claims may exist. Baseline fixtures include duplicate movements, remasters, missing recognition and insufficient usable ratings. None may produce an incorrectly ready release.

A worker test injects duplicate jobs, lease expiry, changed baselines, removed rights and withdrawal while preparation runs. It must produce at most one release per exact version, with obsolete jobs unable to authorize listening. An acquisition test drops acknowledgements, reorders batches, repeats payloads and sends conflicting identifiers. Reconstructed logical events must equal the source fixture, with conflicts explicit and no silent duplication.

The load test runs 50 simulated active sessions for 30 minutes, then 100 for 15 minutes, using the real batch sizes and authorization path. Proposed targets are 95th-percentile committed batch acknowledgement below one second, no lost acknowledged event and error rates below 0.1% excluding deliberately invalid requests. These are acceptance thresholds to measure, not achieved ser-

vice levels. The browser timing assay separately checks audio onset and response timing on representative devices; network performance cannot substitute for it.

A network trace must demonstrate zero MI, GPU or LLM inference calls during a prepared session. A restore drill must recover one complete synthetic study subset with matching object and event hashes, including a withdrawal introduced after the snapshot. An export/replay fixture must reproduce predictions and scores from the same earlier-only information. Permission tests, delivery reconciliation and source/asset licence review complete the release evidence. Failure blocks the affected capability rather than being hidden in an overall “passed” badge.

Pilot and operational release

The proposed 60-registration pilot is included within the total cap and excluded from confirmatory analyses. It measures catalogue coverage, intake effort, task comprehension, timing, missingness, preparation throughput and attrition. It also determines the compact model basis and hyperparameters under the scientific protocol. After pilot review, the team freezes the protocol amendment, catalogue, learner, environment, consent and analysis plan together. Confirmatory participants never enter a moving target of silently updated models or stimulus rules.

Operational dashboards display counts and failures by state, with no need to reveal hidden predictions or personal musical responses. Recruitment is staged against observed readiness throughput and completion burden. Planned pauses and support procedures are part of the participant information. The design does not require a 24-hour availability claim; it requires that interrupted work remain interpretable and recoverable.

Prospective extension to independently measured neural signals

Two measurements, a common event definition

A neuroscience extension would require a new, approved acquisition protocol and independent EEG, MEG or fMRI

observations. The online timestamps cannot serve as a neural ground truth. Physical audio delivery would have to be measured against the acquisition clock, with event markers, device delay, clock drift and dropped-event rules qualified on the actual equipment. Recording identity and exposure order would remain the same scientific objects, while the observation method and temporal resolution would be declared separately.

An illustrative encoding family is

$$b_{iv}(t) = [\mathcal{H}_v * f_{iv}(z_s, c_{ise})](t) + \eta_{iv}(t), \quad (2)$$

where b_{iv} is measured activity, z_s a frozen stimulus representation, c_{ise} admissible context and $\mathcal{H}_v$ the modality-specific observation operation. For fMRI it may include a prespecified haemodynamic model; EEG and MEG require different forward and temporal assumptions. This equation supplies a comparison framework, not a neural implementation of MI. A browser pleasure kernel cannot simply be reused as a haemodynamic or electrophysiological response model.

Human auditory cortex exhibits temporal integration at several scales[9]. That observation motivates explicit temporal support; it does not assign MI windows to cortical areas. A correspondence would have to outperform acoustic and temporal countermodels on independently held-out recordings or participants under matched readout capacity. Regions, time windows, preprocessing, nuisance terms and correction families would be chosen without optimizing on the held-out neural target. A spatially attractive map is not an independent validation set.

Discriminating alternative explanations

Three contrasts are required for an interpretable extension. First, a stimulus representation must add predictive information beyond common acoustic structure and temporal autocorrelation. Second, a participant-conditioned model must improve over the same representation without personal information on matched observations. Third, any association with reported pleasure must be assessed alongside event timing, acoustic salience, motor reporting and the relevant task variables. These contrasts target different alternatives and must not be collapsed into a single brain-model correlation.

The absence of a unique increment is itself informative when features are correlated. If an MI channel and an acoustic baseline span similar predictive subspaces, the experiment may not identify a unique explanatory role even when both predict neural activity. Reliability of repeated measurements constrains the achievable score; a noise ceiling is a property to estimate under stated assumptions, not an assurance that the best-performing model has discovered the mechanism. Out-of-sample neural prediction and causal biological explanation remain different claims[10].

Conditions for a biological interpretation

A manipulation of model code tests a computational dependency. A randomized manipulation of stimulus or context can test its effect on behaviour. Neither, by itself, establishes which neural population or transmitter implements that relation. Specific biological accounts require evidence capable of discriminating them at the biological level. The present online programme therefore uses the terms uncertainty, reward and memory as operationally defined candidate constructs; no anatomical or neurochemical diagnosis is inferred from an audio feature.

A future neural release would link acquisition meta-data, events and derivatives using established neuroimaging conventions[11]. Access restrictions would follow consent and the sensitivity of the observations. The mapping from the behavioural programme to a neural assay would be registered as a new study, with its own sample-size rationale and measurement qualification. It would not be reported as an already implemented component of MI-Lab.

Discussion

MI-Lab treats online listening as a measurement process whose validity depends on stimulus delivery, reporting behaviour and recoverable scientific context. Its proposed contribution is not a new cloud architecture in isolation. It is a set of research objects and acceptance rules that preserve what a participant heard, what they had previously experienced, what a model was allowed to know and how a response entered the analysis. Those distinctions become consequential when learning, temporal dynamics and individual differences are simultaneous targets.

The approach is useful even if MI's present pleasure predictions fail. An immutable stimulus release and explicit exposure history allow investigators to distinguish poor prediction from an incorrectly joined recording, an unqualified timing assumption or a repeated listening event treated as independent. Conversely, correct delivery and complete provenance cannot rescue an invalid behavioural construct. The observation model and the operational record constrain different parts of the inference; both must remain open to criticism.

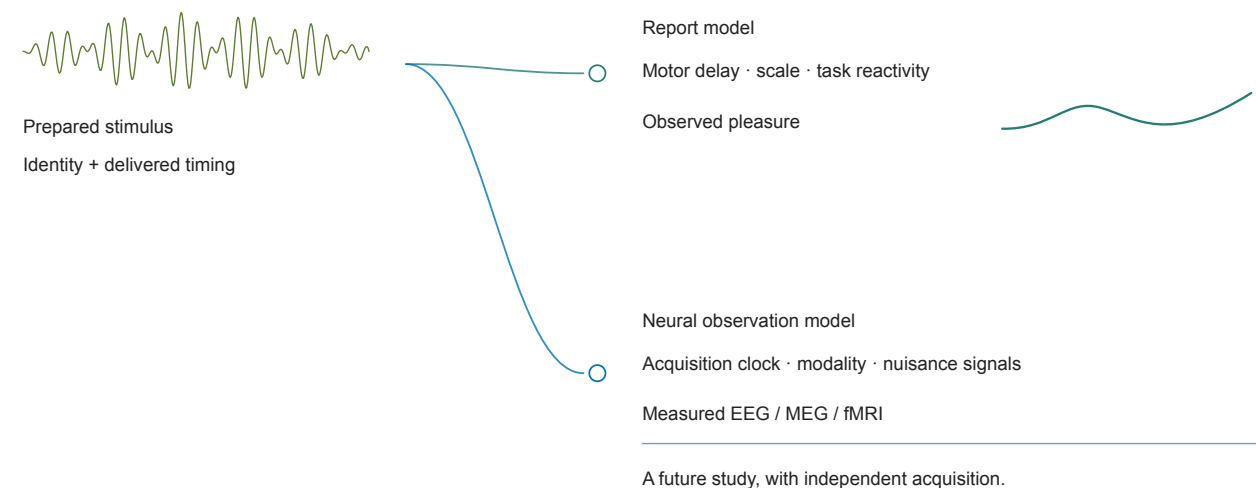
Several limitations are structural. Online audio delivery is heterogeneous, a slider is a delayed and task-dependent report, and the reference catalogue selects a particular listening population. The proposed neural extension adds further acquisition and interpretation requirements. A carefully documented browser session is not a substitute for a qualified EEG, MEG or fMRI experiment. The value of a shared framework is that the extension can retain stimulus and exposure identity while declaring new measurement assumptions, rather than disguising an anatomical label as a recorded signal.

Before use as an empirical methods contribution, the implementation needs measured timing, failure-recovery, privacy and reproducibility results. Acceptance tests below and in Supplementary Methods specify what will count; no deployed performance is claimed. Reporting their failures will be as important as documenting successful operation. A method becomes a research instrument when its precision, failure modes and limits are demonstrated under the conditions in which its outputs are interpreted.

Declarations and research status

This article reports a prospective research design or methodological framework. It contains no newly collected human-participant or neural data. Institutional responsibility, ethics review, participant compensation and operational consent must be established before recruitment. The proposed analyses and acceptance experiments are not reported as completed validations.

Availability. Editable manuscript and vector-figure sources accompany this authoring edition. Restricted audio, participant records and deployment credentials are not included. The current edition has not been submitted or registered by its preparation. Dated architecture



Held-out neural prediction must beat matched acoustic and temporal alternatives.

A brain correlation does not identify a unique biological mechanism.

Prospective acquisition contract · The schematic trace is not neural or participant data.

Fig. 10 | Independent neural observations require a separate assay. A common stimulus release can support behavioural and later neural studies while retaining recording and exposure identity. Reporting and neural observation processes remain distinct. Physical audio synchronization, a modality-specific observation model, nuisance controls and held-out evaluation are prerequisites. The schematic trace is not a measured brain signal.

and earlier protocol records remain distinct from this manuscript.

Author contributions and disclosures. Amaç Erdem is the proposed sole author and originator of Musical Intelligence. This relationship is relevant when evaluating claims about the system. Final CRediT roles, institutional affiliation, funding and competing-interest declarations require the author's verification before submission; none is inferred from the correspondence address. Generative-AI assistance was used in preparing this draft, including editorial development, literature retrieval, figure code and document production. Responsibility for improving the scientific content and the final journal disclosure remains with the author.

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.