

FOUNDATIONAL THEORY PROPOSITION

Relational Mixing

Auditory Existence Before Tools

A mix is a constructed system of auditory relations.

MOODIFY OPEN THEORY / MFY-FT-002

Research Draft / Open Access / Version 0.1

Abstract - This paper proposes a second foundational theory for Moodify: mixing is not fundamentally a collection of plugin operations. It is the construction of auditory relations among sound sources. Equalizers, compressors, reverberators, gain controls, panners, delays, neural renderers, and physical wave-field systems are executors. They alter how sources exist relative to one another, but no individual executor defines the underlying relation. The theory introduces a relational scene model in which source identity, pairwise relations, field, and listening condition jointly determine auditory experience. It separates invariant perceptual intentions such as near/far, wide/narrow, separated/blended, and dominant/supporting from replaceable implementation methods. The resulting architecture gives Moodify a stable protocol layer above changing audio technologies and identifies Judge as the authority that evaluates whether intended relations were realized without violating source identity or evidence boundaries.

Keywords - mixing; auditory relations; sound scene; perceptual attributes; executor independence; wavefield; judgment

1. The Category Error in Tool-Centered Mixing

Audio software presents mixing as a list of operations: EQ, compression, reverb, pan, gain, delay, saturation, and related processors. This organization is practical because software exposes tools as controllable modules. It becomes a theoretical error when the interface taxonomy is mistaken for the object of mixing itself.

Listeners do not ordinarily experience a compressor ratio or an equalizer node as the primary structure of a mix. They experience sources as near or far, front or back, wide or narrow, dense or sparse, dry or wet, bright or dark, stable or moving, separated or blended, dominant or supporting. These terms describe relations and modes of auditory existence, not plugin identities.

The central question is therefore not how to write an equalizer. It is: **How should a sound exist within an auditory space, and how should it relate to every other sound?**

2. The Relational Mixing Proposition

$$E = H(S, R, F, L)$$

$S = \{W1, W2, \dots, Wn\}$ is the set of source signals. R is the organized system of auditory relations among those sources. F is the reproduction field. L is the listener and listening condition. E is the resulting auditory experience. H is the realization relation that converts encoded sources and intended relations into an experienced scene.

Mixing = Construction of R , not accumulation of plugins

R is not a decorative metadata layer. It contains the perceptual organization that makes a collection of tracks become a scene. A useful representation may combine source-level attributes r_i and pairwise or group relations r_{ij} . The model need not assume that every relation is reducible to a single scalar.

3. A Relational Vocabulary

A first relational vocabulary includes the following axes. They are not claimed to be complete or perceptually independent.

Depth: near / far; front / back.

Extent: wide / narrow; focused / diffuse.

Occupancy: dense / sparse; open / crowded.

Environment: dry / wet; direct / reflected.

Spectral impression: bright / dark; soft / hard.

Motion: stable / moving; anchored / drifting.

Separation: separated / blended; distinct / fused.

Hierarchy: dominant / supporting; foreground / background.

These labels are targets for observation, communication, and judgment. They are not fixed knob values. The same target can require different signal operations for different sources, fields, and listener conditions.

4. Executors Are Replaceable

$$X : (S, R, F, L) \rightarrow \text{rendered scene}$$

An executor X is any mechanism that attempts to realize a target relation. Reverb may create perceived distance today. A neural renderer may create comparable distance tomorrow. A physical wave-field synthesis system may later construct distance through an actual spatial field. If all three realize an equivalent intended relation within declared tolerances, the theoretical target remains unchanged while the executor changes.

5. Executor Invariance

Let XA and XB be two different executors. Moodify treats them as functionally substitutable for a declared task when their rendered results preserve required source identity and are judged relationally equivalent within a validated listening scope:

$$\text{Judge}(XA(S, R, F, L), XB(S, R, F, L)) = \text{RELATIONALLY_EQUIVALENT}$$

This is not a claim of waveform identity or bit equivalence. It is a bounded claim that the intended auditory relation has been realized to an accepted perceptual tolerance. The tolerance, listening setup, Judge version, and human escalation rule must be recorded.

Executor invariance is the audio form of a broader architectural principle: **the protocol defines the object; capabilities implement it.** A plugin, model, or renderer is a replaceable provider. The relation schema, evidence contract, and judgment authority remain stable.

6. Relational Optimization

If the desired mix is represented by a target relational state R^* , the optimization problem shifts from choosing favorite processors to selecting an executor and parameters that realize R^* while preserving source identity:

$$(X^*, \theta^*) = \arg \max J(\text{Render}(X\theta, S, R^*, F, L))$$

J must evaluate at least relational fidelity, source preservation, technical validity, uncertainty, and the option to retain the original. A higher score cannot erase a protected voice, instrument, timing relation, or creative intention.

The model also supports non-improvement as a valid result. If no candidate realizes the target relation without unacceptable identity loss, the correct decision is `ORIGINAL`, `HUMAN_REQUIRED`, or `INCONCLUSIVE` rather than forced processing.

7. Consequences for HEAR, JUDGE, and ACT

HEAR estimates the existing auditory scene: source identities, source-to-source relations, uncertainty, and the evidence available for each observation. It describes what exists before recommending an operation.

JUDGE determines which relations serve the work, which must be protected, which are defective or unresolved, and whether a candidate rendering meaningfully improves the scene. It evaluates relations rather than rewarding processor activity.

ACT chooses and controls replaceable executors. Its output is not merely a rendered file. It includes a relation target, execution graph, parameters, evidence, version, and rollback path.

This separation prevents an interface from becoming the theory. Tool lists may change without rewriting Moodify's definition of mixing.

8. Observability and Evidence Boundaries

Some relations can be estimated from a multitrack session or stereo file through level, spectrum, phase, correlation, onset, masking, spatial, and temporal measurements. Others depend on perceptual inference or real-field observation. A file-domain estimate of depth is not automatically a measurement of physical distance. A stereo width metric is not a complete spatial scene.

Every relational claim must identify its evidence source: file-domain measurement, model inference, controlled rendering, real-field measurement, or human listening judgment. Moodify must report uncertainty and refuse false precision where relations are underdetermined.

9. Falsifiability and Research Program

Relational mixing becomes a theory only if its representations can be tested. The research program must ask:

1. Can listeners reliably discriminate and rank proposed relational axes under controlled conditions?
2. Which axes are independent, coupled, context-dependent, or linguistically unstable?
3. Can different executors realize perceptually equivalent relations across multiple fields and listener conditions?
4. Can Judge predict controlled human judgments of relational fidelity and identity preservation?
5. Does relation-first control improve reproducibility, transfer, explanation, and rollback compared with tool-first presets?
6. Which relations require multitrack access, and which can be estimated or altered safely from a stereo mixture?

Failure to obtain reliable discrimination, transfer, or judgment agreement requires revision or rejection of the corresponding relation model.

10. Conclusion

Mixing is the construction of auditory relations among sources. Plugins are execution methods, not the ontology of the mix. Reverb, neural rendering, and wave-field synthesis may all realize perceived distance through different mechanisms. The relation can remain stable while every executor is replaced. This gives Moodify a theory that survives tool turnover: define auditory existence, represent the intended relations, select an executor, and Judge whether those relations were actually realized.

Publication and Open-Theory Statement

This document is a foundational proposition in Moodify Open Theory and carries no individual byline. Its definitions, equations, evidence boundaries, failure conditions, and revisions are open. It reports no completed perceptual experiment and must not be represented as a validated universal theory of mixing.

Data availability. No experimental dataset was created or used.

Ethics declaration. No human subjects, private data, or clinical judgments are involved.

Author contributions. The theoretical direction and core proposition were supplied and approved under human project authority; drafting, structure, typesetting, and consistency checks used an AI-assisted workflow.

Conflict of interest. None declared.

Funding. No external funding declared.

AI-use disclosure. AI assisted with structure, prose, typesetting, and consistency checks. Product and theoretical authority remained human.