Methods

Overview

This project constructs a quantum-inspired generative system for music mashup creation. Beat-aligned audio segments are represented as nodes in a weighted compatibility graph. Quantum and quantum-biological dynamics are used to explore this graph and generate ordered segment sequences, which are subsequently reconstructed into audio.

The pipeline follows:

$$\begin{aligned}&\text{Audio} \to \text{Spectral Features} \to \text{Similarity Graph} \to \text{Hamiltonian}\\&\quad \to \text{Quantum Dynamics} \to \text{Path Extraction} \to \text{Audio Reconstruction.}\end{aligned}$$

1. Audio Segmentation and Database

Music tracks are segmented into short, beat-aligned audio segments (typically 2–4 seconds). Each segment is stored as a standalone WAV file and represented by a `Segment` object containing:

- `id`: unique segment identifier
- `parent_song`: source track name
- `start`, `end`: temporal boundaries in the original audio
- `wav_path`: path to the audio file
- `global_index`: index used for matrix alignment

All segments are stored in a serialized database (`master_db.pkl`), ensuring consistent ordering across feature extraction, graph construction, and quantum evolution.

2. Spectral Feature Construction

All segments share a fixed short-time Fourier transform (STFT) configuration:

- Sampling rate: 22050 Hz
- STFT shape: 1025 frequency bins × 128 frames

From the STFT magnitude, power spectra are computed as:

$$S_{\text{power}} = |\text{STFT}|^2.$$

Mel-spectrogram features are derived from the power spectrum and averaged over time to obtain a fixed-length embedding per segment:

- Feature dimension: $\mathbb{R}^{40}$
- Non-negative values
- One vector per segment

This produces a reproducible and interpretable spectral representation for each audio segment.

3. Feature Normalization and Degeneracy Checks

Each feature vector v is normalized to unit ℓ_2 norm:

$$\hat{v} = \frac{v}{\|v\|}.$$

Explicit checks are applied prior to normalization:

- $\|v\| > 10^{-8}$
- No NaN or infinite values
- Variance collapse across segments is flagged

Segments failing these checks are rejected to prevent similarity degeneracy and graph collapse.

4. Compatibility Score Definition

Pairwise segment similarity is defined using cosine similarity:

$$S_{ij} = \frac{\hat{v}_i \cdot \hat{v}_j}{\|\hat{v}_i\| \, \|\hat{v}_j\|}.$$

To discourage trivial intra-song transitions, a same-song penalty is applied:

$$S_{ij} \leftarrow 0.7 \cdot S_{ij} \quad \text{if segments share the same parent song.}$$

This preserves musical continuity while encouraging cross-song transitions.

5. Graph Construction and Symmetrization

A k-nearest-neighbor graph ($k = 5$–7) is constructed from the similarity matrix. The resulting weighted adjacency matrix A is explicitly symmetrized:

$$A_{ij} = \max(A_{ij}, A_{ji}).$$

This step guarantees that all derived Hamiltonians are Hermitian and admit real eigenvalues. Self-loops are removed, and isolated nodes are verified not to exist.

6. Hamiltonian Definitions

Two Hamiltonians are constructed from the compatibility graph.

6.1 Adjacency Hamiltonian

$$H_A = A.$$

This Hamiltonian directly encodes pairwise similarity but can induce localization around highly connected nodes.

6.2 Graph Laplacian Hamiltonian

$$D_{ii} = \sum_j A_{ij}, \qquad H_L = D - A.$$

The Laplacian Hamiltonian is positive semi-definite and promotes diffusive, exploration-oriented dynamics. Spectral analysis (eigenvalue spread and degeneracy) is used to select the Laplacian Hamiltonian as the primary operator for quantum evolution.

7. Continuous-Time Quantum Walk (CTQW)

Quantum evolution follows the continuous-time Schrödinger equation:

$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle.$$

The initial state is localized at a single segment. Evolution is computed using matrix exponentials or small-step integration. The probability distribution is given by:

$$p_i(t) = |\psi_i(t)|^2.$$

Probability conservation is explicitly enforced at each timestep.

8. Decoherence and ENAQT Modeling

To model environmentally assisted quantum transport (ENAQT), controlled decoherence is introduced by mixing quantum amplitudes with classical noise:

$$\psi \leftarrow (1-\lambda)\psi + \lambda\eta.$$

Here, λ controls noise strength and η is either degree-weighted or uniform noise. Three regimes are evaluated:

- Fully coherent ($\lambda = 0$)
- Intermediate ENAQT regime
- Strongly noisy regime

Each regime produces distinct probability evolution patterns and segment selection behavior.

9. Bio-Inspired Hamiltonian Modulation

A toy biological operator V_{bio} is introduced as a diagonal matrix representing vibrational or energetic modulation:

$$V_{\text{bio}} = \text{diag}(v_1, v_2, \ldots, v_N).$$

The bio-modulated Hamiltonian is defined as:

$$H' = H + \lambda V_{\text{bio}}.$$

Quantum evolution is re-run using H', and results are compared against the baseline Hamiltonian using ℓ_1 distance between probability distributions, Shannon entropy evolution, and path divergence statistics.

10. Quantum-to-Music Path Extraction

Since quantum dynamics produce distributions rather than sequences, a deterministic selection layer is applied. The primary strategy is *argmax with memory*:

- At each timestep, select the most probable unused segment
- Enforce an exclusion window to prevent immediate repeats
- Optional key-continuity constraints may be applied

The result is an ordered sequence of segment identifiers suitable for audio reconstruction.

11. Audio Reconstruction

Selected segments are reconstructed into audio using time-domain stitching:

- Crossfade duration: 50–100 ms
- Linear fade-in and fade-out
- Final normalization to prevent clipping

This produces the final mashup audio.

12. Spectral Crossfade Validation (Analysis Only)

An optional analysis compares time-domain crossfading with frequency-domain (STFT magnitude) crossfading. While spectral blending preserves harmonic continuity, audible improvements are marginal. For robustness, time-domain crossfading is used in the final pipeline, while spectral crossfading is retained as validation of frequency-domain operation.

Summary

This methodology establishes a reproducible signal-to-graph-to-quantum-dynamics-to-audio synthesis pipeline. Quantum and quantum-biological effects influence *selection*, while digital signal processing governs *realization*, maintaining a clean separation of concerns suitable for research extension.