

ChordyV: A Symbolic Harmonic Progression and Tonal Audit System for Audio Key Detection

Vie (Mulfie Ansyori Sofyan)

Independent Researcher / ChordyV Developer

masprodios.chordyv@gmail.com

ABSTRACT

This paper presents the submission of ChordyV to the MIREX (Music Information Retrieval Evaluation eXchange) Audio Key Detection task. Unlike conventional black-box models that predict global musical keys directly from raw spectrograms—often misled by percussive transients, unpitched noise, and dynamic variations—ChordyV introduces a deterministic, multi-stage hybrid pipeline. The system isolates musical stems using deep source separation, locks onto structural bar grids, tracks harmonic chord sequences via dual-layer Constant-Q Transform (CQT) arrays, and refines them through passing-bass elimination. Global tonality is resolved via the ChordyV Tonal Auditor, a symbolic progression engine that evaluates 24 major and minor keys using diatonic pitch-class distributions, cadential progression rules ($V \rightarrow I$, $ii \rightarrow V$), and intro pedal-bass analysis. The output strictly complies with MIREX single-line tab-delimited ASCII formatting and satisfies fully offline containerized execution.

1. INTRODUCTION

Global audio key detection aims to determine the primary tonic pitch and mode (major or minor) of an audio track. Standard frame-level neural classifiers often suffer from confusion between relative keys (e.g., C major vs. A minor) or parallel modes due to heavy drum layers, non-harmonic timbres, and complex polyphony.

ChordyV addresses this challenge by framing key identification through a music-theoretically grounded, symbolic paradigm: **extracting authentic chord progressions first, then resolving global tonality**. By combining stem separation, metric downbeat locking, harmonic progression analysis, and cadential validation, ChordyV replicates the analytical reasoning of a trained musician.

2. SYSTEM ARCHITECTURE

The ChordyV key detection pipeline operates sequentially through five distinct stages:

$\text{Raw Audio} \rightarrow \text{Stem Shaping} \rightarrow \text{Metric Locking} \rightarrow \text{Dual-CQT \& Chord Extraction} \rightarrow \text{Tonal Audit \& Key Output}$

2.1. Neural Stem Ingestion & Spectral Shaping

The framework isolates fundamental acoustic roles using deep source separation (htdemucs_6s):

- **Root Subsystem:** The extracted bass stem is combined with low-level accompaniment and processed through a Butterworth low-pass filter (cutoff 260 Hz), attenuating melodic bleed to isolate clean fundamental bass contours.
- **Harmonic Subsystem:** Polyphonic instrumental stems (keyboards, guitars) undergo complementary high-pass filtering (cutoff 130 Hz), preventing bass energy from interfering with upper triad identification.

2.2. Metric Induction and Phase Alignment

Temporal profiling is conducted via multi-feature autocorrelation combining Harmonic Change Detection Functions (HCDF), low-frequency spectral flux, and transient onset envelopes:

- **Meter Classification:** The system dynamically classifies binary (4/4) and compound/ternary (6/8) time signatures.
- **Downbeat Locking:** The downbeat anchor (t_0) is resolved by optimizing transient coincidence across onset vectors, establishing a stable bar grid for subsequent harmonic extraction.

2.3. Dual-Layer CQT Feature Extraction & Literal Chord Mapping

Pitch tracking operates over two acoustic registers:

- **Fundamental Bass Register:** Lower-octave logarithmic CQT arrays compute fundamental bass distributions per beat.
- **Upper Harmonic Register:** Mid-to-high CQT arrays capture upper triad textures, matched against prototypical major, minor, suspended, and diminished templates.

2.4. Passing-Bass Suppression & Structural Chord Extraction

- **Passing-Note Suppression:** Weak-subdivision bass changes maintaining high spectral cosine similarity (≥ 0.72) are treated as linear passing motions and merged, preventing false harmonic fragmentation.
- **Functional Pop Dynamic Translation:** Inversions and slash bass intervals are parsed into structured chord representations per bar.

2.5. Tonal Auditor & MIREX Key Formatting

Once the bar-level chord sequence is formed, the **ChordyV Tonal Auditor** resolves the primary tonality across all 24 musical keys (12 major and 12 minor):

- **Intro Pedal-Bass Tracking:** The system monitors bass pitch-class occurrences across the initial 16 bars. A dominant bass pedal ($\geq 35\%$ presence) injects anchor weight into corresponding tonal candidates.
- **Diatonic Degree Weighting (L_1):** Accumulated chord roots are evaluated against diatonic scale degrees (tonic, dominant, subdominant, and mediant), while out-of-scale roots receive negative penalties.
- **Cadential Progression Scoring (L_2):** Sequential bar transitions are evaluated for strong cadential movements, prioritizing $V \rightarrow I$ (with downbeat bonuses), $ii \rightarrow V$, and $iv \rightarrow I$ in both major and minor modes.
- **Probabilistic Scoring:** Normalized L_1 and L_2 matrices are fused through an exponential distribution function to select the winning tonic and mode.
- **MIREX Output Formatting:** The predicted key is written as a single-line tab-delimited ASCII file:

$\angle \text{tonic} \backslash \text{mode} \backslash \text{n}$

(e.g., $C \backslash \text{major}$ or $G \# \backslash \text{minor}$).

3. RUNTIME AND COMPLIANCE SPECIFICATIONS

- **Offline Execution:** Pre-trained model weights are pre-cached during the build step (`TORCH_HOME=/app/temp/chordyv_cache`), guaranteeing deterministic inference inside network-restricted MIREX evaluation environments.
- **Command Line Interface:** The system natively supports both positional (`python3 chordyv_key.py %input %output`) and flagged CLI formats (`-i %input -o %output`).
- **Resource Safety:** Ephemeral stem files and intermediate buffers are automatically purged upon key generation to maintain a minimal disk footprint.

4. CONCLUSION

ChordyV bridges deep audio representation and symbolic harmonic theory. By determining global key from clean, metrically aligned chord progressions and cadential audit logic, ChordyV offers an interpretable and structurally sound architecture for the MIREX 2026 Audio Key Detection task.

REFERENCES

1. E. Gómez, "Tonal description of polyphonic audio for music information retrieval," *INFORMS Journal on Computing*, vol. 18, no. 3, pp. 294–304, 2006.
2. C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*, Oxford University Press, 1990.
3. A. Défossez, N. Usunier, L. Bottou, and F. Bach, "Music Source Separation in the Waveform Domain," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 29, pp. 642–652, 2021.
4. C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. P. W. Ellis, "mir_eval: A Transparent Implementation of Common MIR Metrics," in *Proc. 15th Int. Conf. on Music Information Retrieval (ISMIR)*, 2014.