Automatic Guitar Chord Transcription

***Abstract*—This project focuses on automatic guitar chord transcription from audio signals. The objective is to design and compare a classical signal-processing pipeline with a modern deep-learning-based system. Both approaches are evaluated on real musical recordings using the Isophonics Beatles dataset.**

I. Goals

The objective of this laboratory assignment is to design, implement, and evaluate a system capable of automatically transcribing guitar chord progressions directly from audio recordings. The task involves detecting chord boundaries over time and assigning the correct chord label to each segment of the audio.

The project focuses on two complementary approaches:

- **A classical signal-processing pipeline**, based on:
  - harmonic/percussive source separation (HPSS),
  - chroma feature extraction,
  - template matching for major/minor triads,
  - temporal smoothing using filtering or post-processing.
- **A deep learning model**, trained to perform chord classification directly from spectrogram- or chroma-based input representations.

The final goal is to compare the performance, robustness, and limitations of these two approaches, and to analyze how they behave on real musical recordings with varying harmonic complexity.

II. Related Work

Automatic chord recognition has been widely explored in the MIR community, with two dominant families of approaches: classical signal-processing pipelines and deep-learning-based methods. The objective of this project is to implement one representative technique from each family and compare their behaviour on real audio.

A. Classical Approaches

Recent classical systems rely on chroma features, template matching and temporal smoothing. Mauch et al. [1] showed that applying harmonic/percussive source separation and using robust chroma representations significantly improves the reliability of template-based chord estimation. Similarly, Ni et al. [2] proposed a refined template-matching pipeline using normalized major/minor templates, cosine similarity and smoothing. The classical baseline developed in this project follows these principles through HPSS, CQT-based chroma extraction and a normalized template-matching stage.

B. Deep Learning Approaches

More recent work has focused on learning chord representations directly from audio. Korzeniowski and Widmer [3] introduced fully convolutional networks trained on spectrograms, achieving state-of-the-art performance without hand-engineered features. Chen and Su [4] further improved results by combining CNN layers with recurrent units to capture temporal dependencies. These architectures inspire the deep-learning model planned in this project, based on spectrogram inputs and either convolutional or recurrent processing. The Isophonics Beatles annotations [6] are used in this work to ensure comparability with these prior systems.

III. Progress

During the first part of the laboratory assignment, I implemented a complete classical chord-recognition pipeline, from audio preprocessing to evaluation on annotated Beatles recordings.

A. Preprocessing and Feature Extraction

The preprocessing stage consists of mono audio loading at 22.05 kHz, followed by Harmonic/Percussive Source Separation (HPSS) to isolate harmonic content. Chroma CQT features were then extracted using `librosa`.

Fig. 1 shows the waveform of *Let It Be*, while Fig. 2 presents the corresponding CQT-based chromagram. These visualizations were used to verify the correctness of the pipeline and the effectiveness of HPSS in reinforcing harmonic clarity.

A first challenge in this stage was ensuring that the chroma vectors were consistently normalized and that HPSS provided meaningful improvements for harmonic analysis.

B. Classical Template-Matching Baseline

A full template-matching chord recognizer was developed. The method uses 24 normalized major/minor triad templates and computes cosine similarity between each template and every chroma frame. The highest-scoring template gives the framewise chord prediction. A temporal mode filter smooths short-term fluctuations, and the final predictions are converted into time-based chord segments.

The resulting chord timeline for *Let It Be* is shown in Fig. 3. One challenge here was the high sensitivity of chroma features to overtones and noise; temporal smoothing significantly improved stability and interpretability.

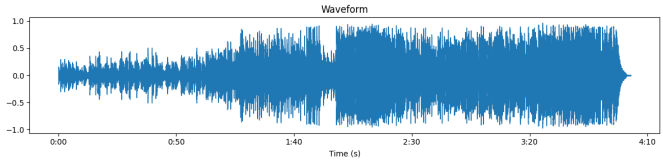

Fig. 1. Waveform of *Let It Be*.

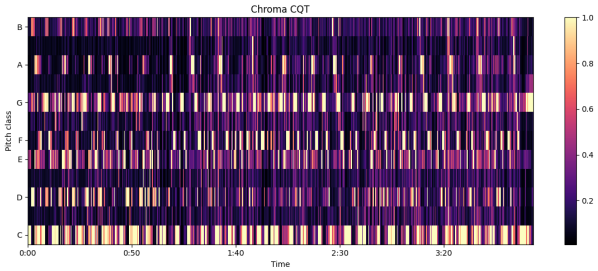

Fig. 2. CQT-based chromagram for *Let It Be*.

C. Ground Truth Parsing and Label Harmonization

The Beatles `.lab` annotation files contain complex symbolic chords such as `A:min/b7`, `F:maj7`, or `C/7`, which are incompatible with a triad-only model. I implemented a parser and harmonization function to map these labels to a reduced 24-class major/minor vocabulary.

This required handling slash chords, chord extensions, and ambiguous cases while preserving the root and major/minor quality. This harmonization step was essential for fair comparison between predictions and ground truth.

D. Alignment of Predictions and Ground Truth

To evaluate the system, each chroma frame time was aligned with the ground-truth segment that contains it. This produced a frame-level ground-truth sequence matching the length of the predictions. Ensuring correct temporal alignment was a key step, as any mismatch would distort the accuracy measurement.

E. Initial Results

Frame-level accuracy was computed on two Beatles tracks.

For *Let It Be*, the classical baseline achieved an accuracy of **69.97%**. This relatively high value is explained by the song's simple harmonic progression (C–G–Am–F), on which template matching performs well.

For *I'll Follow the Sun*, accuracy dropped to **20.52%**. This track contains more frequent chord changes and modulations, clearly exposing the limitations of a simple template-matching system.

These results confirm expected behavior: the classical method performs well on stable, diatonic material, but struggles with more complex harmonic structures.

IV. Futur Work

With the classical baseline and evaluation framework completed, the next phase of the project focuses on implementing a deep learning model for chord recognition and comparing it with the signal-processing approach.

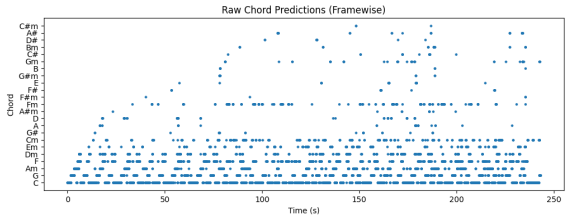

Fig. 3. Classical template-matching chord timeline for *Let It Be*.

The model will be based on a CNN or CRNN architecture trained on audio-derived features such as log-mel spectrograms or CQT chroma patches. Training examples will be constructed from fixed-size windows paired with the harmonized major/minor labels used in the classical baseline. Convolutional layers will be used to learn local harmonic patterns, and recurrent layers may be added to capture temporal dependencies. The model will output framewise chord probabilities over 24 classes.

Training will rely on a subset of the Isophonics Beatles dataset with validation splits and early stopping. Data augmentation techniques such as pitch shifting may be used if beneficial. After training, predictions will be smoothed and segmented using the same procedure as the classical method to enable a fair comparison.

The deep-learning system will then be evaluated with the existing alignment and frame-level accuracy metrics on the same tracks, including *Let It Be* and *I'll Follow the Sun*. The comparison will examine improvements in accuracy and robustness, especially on passages where template matching is limited. Qualitative inspection of timelines will complement the quantitative evaluation.

This final step will provide a clear comparison between classical and learned approaches and assess their respective performance on real musical material.

V. Conclusion

This intermediate report presented the implementation and evaluation of a classical chord-recognition baseline based on chroma features and template matching. The system performs well on simple harmonic material, as illustrated by the results on *Let It Be*, but shows clear limitations on more complex songs such as *I'll Follow the Sun*. These findings highlight the need for a more flexible, data-driven approach. The next stage of the project will therefore focus on developing a deep learning model and conducting a systematic comparison between the two methods to assess their respective strengths and weaknesses in automatic chord transcription.

References

[1] M. Mauch, M. Davies, and S. Dixon, "Improving Chord Recognition: HPCP Features and Harmonic/Percussive Separation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[2] Y. Ni, M. McVicar, R. De Bie, and S. Dixon, "An Improved Chord Recognition Pipeline Using Harmonic Template Matching and Post-Processing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
[3] F. Korzeniowski and G. Widmer, "A Fully Convolutional Deep Network for Chord Recognition," in *Proc. IEEE International Conference on Music Information Retrieval (ISMIR)*, 2016.
[4] K.-H. Chen and L. Su, "Chord Recognition Using Convolutional Neural Networks with Bidirectional Recurrent Neural Layers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.
[5] C.-W. Wu and A. Lerch, "Automatic Chord Recognition Using Data Augmentation and Transfer Learning," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2020.
[6] M. Mauch and S. Dixon, "Simultaneous Estimation of Chords and Musical Context Using Probabilistic Models," in *Proc. International Society for Music Information Retrieval Conference (ISMIR)*, 2010.