Symbolic and Audio Key Detection Based on a Hidden Markov Model

Néstor Nápoles
McGill University, CIRMMT
nestor.napoleslopez@mail.mcgill.ca

Claire Arthur
Georgia Institute of Technology
claire.arthur@gatech.edu

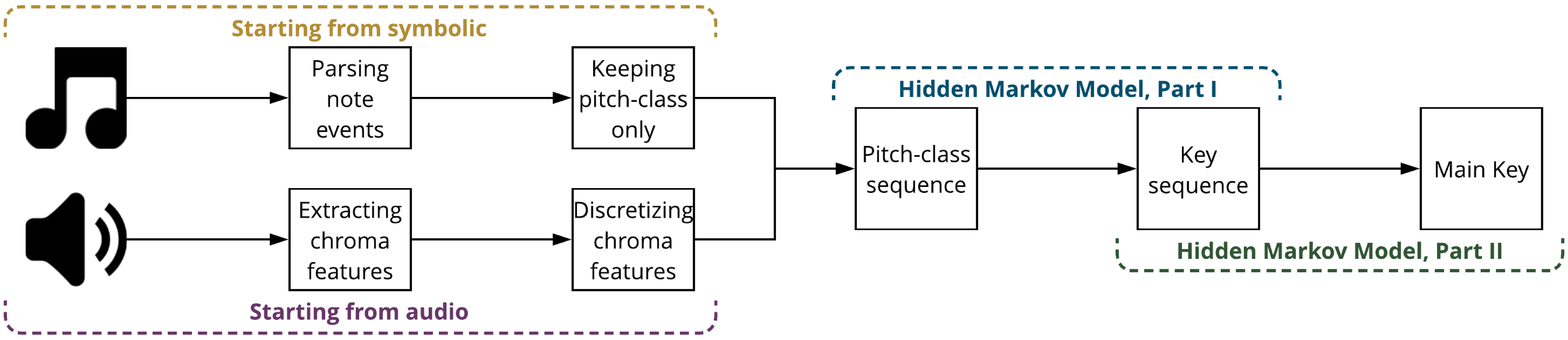
Ichiro Fujinaga
McGill University, CIRMMT
ichiro.fujinaga@mcgill.ca

McGill

Centre for Interdisciplinary Research in Music Media and Technology

Summary

This project started as a symbolic key detection algorithm. In order to participate in the Music Information Retrieval EXchange (MIREX) key detection task, additional audio processing stages have been included. The user input—whether it is a symbolic or an audio file—is pre-processed to obtain a sequence of pitch classes, which becomes the input of the Hidden Markov Model (HMM). The HMM is divided in two parts, the first part outputs a sequence of temporary keys (i.e., tonicizations) and the second part outputs a main key. Both parts rely on a geometric model of key distance (for transition probabilities) and key-profiles (for emission probabilities). The model has been evaluated in both its symbolic and audio implementations.



Starting from symbolic

If the input of the model is a symbolic file, converting a sequence of notes into pitch classes is a trivial process (e.g., performing a modulo operation over MIDI note numbers).

Starting from audio

In the case of receiving audio input, the pitch-class sequence is obtained by discretizing a sequence of chromagram features. The chromagram features have been extracted using the *NNLS Chroma* method [1]. An algorithm iterates over each audio frame and determines which chroma components should be included in the sequence of pitch classes.

Hidden Markov Model, Part I

The first part of the HMM maps a sequence of pitch-classes into a sequence of temporary keys. A simple intuition of this analysis is: **for every pitch-class in a sequence of pitch classes, what is the key that most likely generated this pitch-class?**

Hidden Markov Model, Part II

The second part of the HMM analyzes each of the temporary keys generated in the previous part and outputs a single (i.e., main) key. A simple intuition of this analysis is: **for a sequence of temporary keys in a musical fragment, what is the key that better explains the sequence?**

Results

The model has been tested with two datasets, one for symbolic input and one for audio input. The first dataset consists of 124 MIDI files of music pieces from the Baroque period to Post-romanticism. The second dataset consists of 260 audio files: 164 audio files synthesized from MIDI files of classical music pieces with varied instrumentation, 45 audio recordings of classical music pieces, and 51 popular pieces taken from the McGill Billboard corpus [2]. The evaluation metrics followed are the ones from the key detection task in MIREX.

- **Symbolic**: 124 MIDI files with 87.6% accuracy
- **Audio**: 260 audio files with 84.8% accuracy

The implementation of this model has been submitted to MIREX in 2018. The results from its performance will be included in the future.

Acknowledgments

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

Fonds de recherche Société et culture Québec

Social Sciences and Humanities Research Council of Canada — Conseil de recherches en sciences humaines du Canada

Canada

References

[1] Matthias Mauch and Simon Dixon. Approximate note transcription for the improved identification of difficult chords. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[2] John Ashley Burgoyne, Jonathan Wild, and Ichiro Fujinaga. An expert ground truth set for audio chord recognition and music analysis. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 633–638, 2011.

Presented at ISMIR 2018, Paris, France, September 23–27, 2018