DMRN+15: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2020

QUEEN MARY UNIVERSITY OF LONDON
TUE 15 DECEMBER 2020

A Modular System for Harmonic Structure Analysis of Music

Andrew McLeod and Martin Rohrmeier*

Digital and Cognitive Musicology Lab, EPFL, Switzerland, andrew.mcleod@epfl.ch

***Abstract*— Harmonic structure analysis is the task of labeling an input musical piece (be it a score, MIDI, or audio) with chord and local key information. The task involves many interconnected dependencies at various levels of granularity (from low to high: frames, notes, chords, and keys). In this work, we propose a system with a modular design, allowing each component to regard the data at the appropriate level.**

***Index Terms*—** Chord, key, harmonic analysis

I. TASK AND SYSTEM

Previous work on full harmonic analysis [1, 2] has treated the input as a sequence of input frames, assigning a label to each with various (sequential and non-sequential) neural network architectures. Our modular system, on the other hand, models each aspect of the analysis at its corresponding level of granularity. We hypothesize that such a design will allow our system to be more interpretable, as well as more adaptable to various use cases.

We use a very large vocabulary of chords and keys taking on a full characterization as used in music theory. Chord roots and key tonics may be any pitch A–G, double-flat to double-sharp (35 total). Chords may be major, minor, augmented (each with no, major, or minor 7th), or diminished (with no, minor, or diminished 7th) (12 total); in any inversion (3 for triads, 4 for 7th chords). This totals 1540 chords and 70 keys (major or minor for each tonic).

Our system is composed of 6 modules, each with a well-defined input and output (see Fig. 1). Its input can be a musical score (notes), a MIDI file (notes), or frames of an audio spectrogram, though we currently use only musical scores, and its output is a list of (absolute or relative) chord symbols and local keys, each corresponding to a range of inputs.

The *Chord Transition Model* (CTM) takes the system's input vectors and outputs the probability of each being the start of a new chord. The *Chord Classification Model* (CCM) takes as input a list of input vectors belonging to the same chord, and outputs a distribution over all chords. The *Chord Sequence Model* takes as input a sequence of chords (whose root pitch is relative to the current key's tonic), and outputs a distribution over the next chord at each step. The *Key Transition Model* takes as input a sequence of relative chords and outputs the probability of each being the start of a new local key. The *Key Sequence Model* takes as input the sequence relative chords from the previous local key section of a piece, plus the first chord symbol of a new key section (still relative to the previous tonic), and outputs a distribution over the next local key (as an interval from the previous tonic plus major or minor). The *Initial Chord Model* outputs a distribution over the first relative chord symbol of a piece given the key.

*Research supported through the Swiss National Science Foundation within the project "Distant Listening – The Development of Harmony over Three Centuries (1700–2000)" (Grant no. 182811).

Table 1: Evaluation results.

Chord			Key		
Root+Triad	+7ths	**+Inv**	Tonic	**+Mode**	**Full**
0.47	0.32	0.26	0.45	0.34	0.15

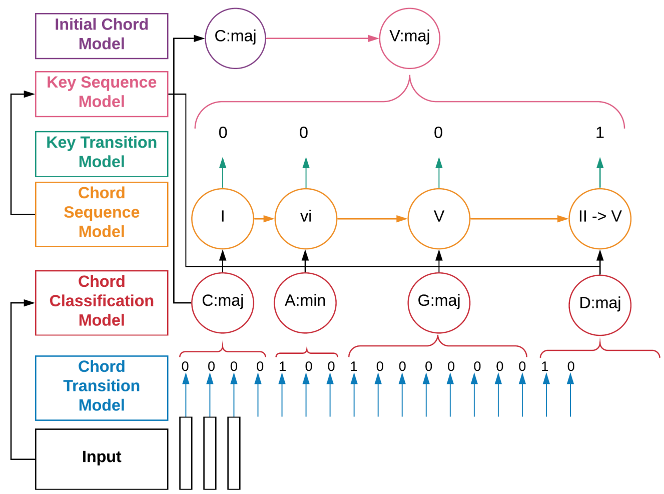

Figure 1: Overview of our fully integrated system. Each component depends on the component directly below it, in addition to the arrows.

II. RESULTS AND DISCUSSION

We train and evaluate the system on a private set of annotated musical scores from a variety of composers (including the Annotated Beethoven Corpus [3]). Results are shown in Table 1. Each value is the average proportion of each piece with the correct label. Although each of our modules is currently very simple (most are a single-layer LSTM with a softmax), our results are promising.

Our system's modular allows us to train and improve each component independently, treating each as a black box. Noisy training methods such as scheduled sampling could also be used to make our model more robust to decoding errors. In future work, we plan to adapt the system to different input formats (MIDI and audio—only the CTM and CCM would need to be re-trained), and use the system in a human-in-the-loop way for annotation where a human can force the search process to go through particular manually-input labels.

III. REFERENCES

[1] T.-P. Chen and L. Su, "Harmony transformer: Incorporating chord segmentation into harmony recognition," in *ISMIR*, nov 2019.

[2] G. Micchi, M. Gotham, and M. Giraud, "Not all roads lead to rome: Pitch representation and model architecture for automatic harmonic analysis," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, pp. 42–54, may 2020.

[3] M. Neuwirth, D. Harasim, F. C. Moss, and M. Rohrmeier, "The annotated beethoven corpus (ABC): A dataset of harmonic analyses of all beethoven string quartets," *Frontiers in Digital Humanities*, 2018.