CHORD DETECTION METHOD FOR MIREX 2008

Matti Ryynänen and Anssi Klapuri
Department of Signal Processing, Tampere University of Technology, Finland
matti.ryynanen@tut.fi, anssi.klapuri@tut.fi

ABSTRACT

We submit a method for the MIREX 2008 task "Audio Chord Detection". The method uses a 24-state HMM consisting of 12 states for both major and minor triads. The observation likelihoods are computed by mapping the pitch saliences into a pitch-class representation and comparing them with trained profiles for major and minor chords. Probabilities of between-chord transitions are estimated from training data and Viterbi decoding is used to find a path through the chord models.

1 INTRODUCTION

The chord transcription problem has recently become a topic of interest for several researchers, as a part of automated analysis of harmonic content in music signals. This has resulted in numerous publications on the topic, including [9, 1, 3, 2, 7, 5]. The most popular approach is to extract low-level features describing harmonic content, e.g., chroma or pitch class profile (PCP) vectors, and then use a hidden Markov model (HMM) with states corresponding to different chords to perform the recognition. Knowledge of typical chord progressions can be applied as well.

Our submission follows this common approach and uses PCPs and a chord-HMM for the chord transcription task. The submitted method has been published in [8] and is briefly introduced in the following. For details, please see [8].

2 METHOD DESCRIPTION

The method begins with pitch salience estimation where salience, or strength, of each fundamental frequency (F0) is calculated as a weighted sum of the amplitudes of its harmonic partials in a spectrally whitened signal frame. The salience function is evaluated in overlapping 92.9 ms frames, with 23.2 ms interval between successive frames. The calculations are similar to [4].

The salience function values are then mapped into a pitch-class representation, where notes in different octaves are considered equivalent. Instead of one PCP vector, we use two pitch-class profiles: one for low-register MIDI notes 26–49 and one for high-register MIDI notes 50–73. The low and high-register profiles for major and minor triads are

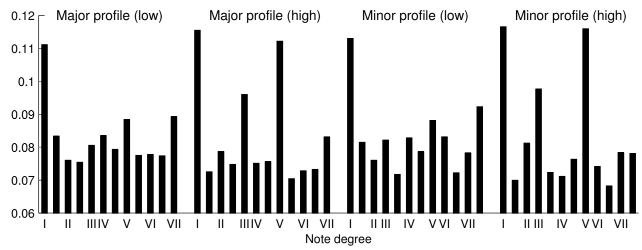

Figure 1. Major and minor chord profiles for low and high registers.

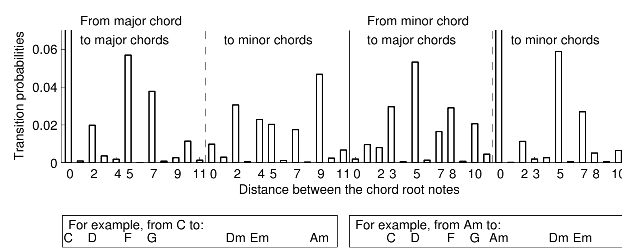

Figure 2. The estimated chord transition probabilities from major and minor chords. The text boxes show examples for transitions from C major chord and A minor chord.

estimated from the first eight albums by the Beatles with annotations provided by Harte and colleagues. Instead of estimating chord profiles for all 24 major/minor triads, we estimate the profiles only for major and minor chord types. This circumvents the problem of insufficient training data for particular chords (e.g., F♯ major in the Beatles data, see [6] for chord distributions).

Figure 1 shows the estimated low and high-register profiles for both major and minor chords. The low-register profile captures the bass notes contributing to the chord root whereas the high-register profile has more clear peaks also for the major or minor third and the fifth.

We also estimate a chord-transition bigram from the Beatles data. The transitions are independent of the key so that only the chord type and the distance between the chord roots matters. Figure 2 illustrates the estimated chord-transition probabilities. The probability to stay in the chord itself is a

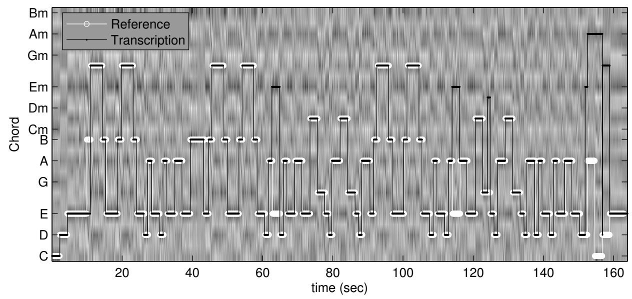

Figure 3. Chord transcription of "With a Little Help From My Friends" by The Beatles. Observation likelihoods are shown for all major and minor triads. The reference chords are indicated by the white line and the transcription by the black line.

free parameter which controls the amount of chord changes.

Finally, we define a chord HMM with 24 states, twelve states for both major and minor triads. The observation likelihoods for each chord are calculated by comparing the low and high-register profiles with the estimated trained profiles shown in Fig. 1. Then Viterbi decoding through the chord HMM produces chord labeling for each analysis frame. Currently our method does not take silent or no-chord segments into account but outputs a chord label in each frame. Figure 3 shows an example transcription for "With a Little Help From My Friends" by The Beatles.

3 ABOUT THE IMPLEMENTATION

The method has been implemented with C++ and it takes less than four seconds to process one minute of audio on a 3.2 GHz Pentium 4 processor.

4 RESULTS AND DISCUSSION

We participated to the first subtask where submitted methods were pretrained with desired material or otherwise fixed. The best results were obtained by the submission of Bello & Pickens (66% correct). Due to the rather similar approaches (namely using HMMs) to the task, it is not surprising that differences in the results are very small (e.g., Mehnert 65%, Ryynänen & Klapuri 64%, Papadopoulos & Peeters 63%, Khadkevich & Omologo 63%). As we already noted in [8], our method tends to confuse major and minor modes. When these errors are ignored, the submitted method performs as well as the method by Bello & Pickens (both 69% correct). The submitted method was among the fastest submissions (the method by K. Lee was clearly the fastest).

In the second subtask, the methods were run in three-fold cross-validation (train with 2/3 and test with 1/3 of the database). The method by Uchiyama, Miyamoto, and Sagayama (72% correct) clearly outperformed the other submissions.

5 ACKNOWLEDGMENTS

The authors would like to thank the IMIRSEL group for running the evaluations and organizing MIREX in general, the task organizers, people involved in the discussions, and all the participants.

6 REFERENCES

[1] J. P. Bello and J. Pickens. A robust mid-level representation for harmonic content in music signals. In *Proc. 6th International Conference on Music Information Retrieval*, pages 304–311, 2005.

[2] E. Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[3] C. A. Harte and M. B. Sandler. Automatic chord identification using a quantised chromagram. In *Proc. 118th Audio Engineering Society's Convention*, 2005.

[4] A. Klapuri. Multiple fundamental frequency estimation by summing harmonic amplitudes. In *Proc. 7th International Conference on Music Information Retrieval*, pages 216–221, 2006.

[5] K. Lee. *A System for Acoustic Chord Transcription and Key Extraction from Audio Using Hidden Markov Models Trained on Synthesized Audio*. PhD thesis, Stanford University, Mar. 2008.

[6] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):291–301, Feb. 2008.

[7] H. Papadopoulos and G. Peeters. Large-scale study of chord estimation algorithms based on chroma representation and HMM. In *Proc. International Workshop on Content-Based Multimedia Indexing*, pages 53–60, June 2007.

[8] M. P. Ryynänen and A. P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32(3):72–86, 2008.

[9] A. Sheh and D. P. Ellis. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proc. 4th International Conference on Music Information Retrieval*, 2003.