CHORD ESTIMATION USING CHORD TEMPLATES AND HMM

Hélène Papadopoulos
IRCAM / CNRS-STMS
Sound Analysis/Synthesis Team, Paris - France
papadopo@ircam.fr

Geoffroy Peeters
IRCAM / CNRS-STMS
Sound Analysis/Synthesis Team, Paris - France
peeters@ircam.fr

ABSTRACT

This document describes a submission to the Music Information Retrieval Evaluation eXchange in the Audio Chord Recognition task. A set of chroma vectors representing the pitch content of the audio file over time is extracted. From these observations the chord progression is then estimated using hidden Markov models. The system takes into account the presence of higher harmonics of pitch notes and includes some music knowledge. In this approach, no training is needed.

1 INTRODUCTION

In Western tonal music, the chord progression determines the harmonic structure of a piece of music. Analysis of chord progression therefore plays a crucial role in the understanding of this music. Automatic chord recognition has become a major field of MIR and many approaches have already been proposed. The MIREX 2008 audio chord detection task is a first step to define an accurate methodology in order to make the comparison of the results possible. Details about the system we present here can be found in [1] and [2].

2 FEATURE EXTRACTION

The front-end of our system is based on the extraction of a set of feature vectors (*chroma vectors* [3]) that represent the audio signal. The audio signal is converted to mono. We only consider frequencies above $61.7Hz$, which corresponds to a B1 midi note. The upper frequency limit is set to $1kHz$. Since the instruments may have been tuned according to a reference pitch different from the standard $A4 = 440Hz$, it is necessary to estimate the tuning of the track. Here, the tuning is estimated using the method proposed in [4]. The sequence of chroma vectors over time is known as chromagram. Existing methods to compute a chromagram present some variances but follow in general two steps: first a semitone pitch spectrum is either computed from the FFT or obtained by the Constant Q Transform (CQT [5]) because the center frequencies of the CQT are spaced according to the frequencies of the equal-tempered scale; then the semitone pitch spectrum is mapped to the chroma vectors. As proposed in [6], smoothing the semitone pitch spectrum provides a reduction of transients and noise in the signal. We obtain a sequence of 12-dimensional vectors that are suitable feature vectors for our analysis.

3 MODEL

We consider an ergodic 24-states HMM, each state representing a single chord. The chord lexicon is composed of 24 Major and minor triads (C Major, C# Major, ..., B Major, C minor, ..., B minor). Each state in the model generates an observation vector, the chroma feature, with some probability. This is defined by the **observation probabilities**. The transitions between chords result from musical rules that should be reflected in the **state transition matrix**. The state-transition matrix we use is based on cognitive experiments. Given the observations, we estimate the most likely chord sequence over time in a maximum likelihood sense.

3.1 Initial state distribution

Since we do not know *a priori* which chord the piece begins with, the initial state distribution is uniformly initialized at $\frac{1}{24}$ for each of the 24 states.

3.2 Observation symbol probability distribution

The probabilities $P(c_i(t_m)|\mathbf{O}(t_m))$ are obtained by computing the correlation between the observation vectors (the chroma vectors) and a set of chord templates which are the theoretical chroma vectors corresponding to the $I = 24$ major and minor triads. Each chord template is a 12-dimensional vector which contains the theoretical amplitude values of the notes and their harmonics composing a specific chord. The chord templates are constructed considering the presence of the higher harmonics of the theoretical notes it consists of, relying on the model presented in [7]: the amplitude contribution of the h^{th} harmonic composing the spectrum of a note is set to 0.6^{h-1}.

Chord symbol probabilities computation: At each time instant t_m, we compute the correlation between the observation vector $\mathbf{O}(t_m)$ and the 24 chord templates $\mathbf{CT}_i, i \in [1, 24]$.

$$For \quad i = 1 \dots 24, \quad P(\mathbf{O}(t_m)|c_i(t_m)) = <\mathbf{O}(t_m), \mathbf{CT}_i> \tag{1}$$

The 24 values $P(\mathbf{O}(t_m)|c_i(t_m))$ are normalized so that

$$\sum_i P(\mathbf{O}(t_m)|c_i(t_m)) = 1 \tag{2}$$

3.3 State transition probability distribution

The chord transition matrix is obtained using values corresponding to correlations between key profiles obtained from perceptual tests by Krumhansl. These correlations were first used by [8] to derive a key transition matrix.

3.4 Chord progression over time detection

The optimal succession of chords over time is found using the Viterbi decoding algorithm [9] which gives us the most likely path trough the HMM states given our sequence of chroma observations.

4 ANALYSIS OF THE RESULTS

4.1 Task description

The mirex 2008 Audio Chord Detection task was divided into two subtasks:

- in the first one the systems were pretrained and they were tested against 176 Beatles songs
- in the second one systems were trained on 2/3 of the Beatles dataset and tested on 1/3

Our system does not need any training, we thus participated to the first subtask.

4.2 Evaluation measure

Overlap score was calculated as the ratio between the overlap of the ground truth and detected chords and ground truth duration. A secondary score was calculated by ignoring the major-minor variations of the detected chord (e.g., C major == C minor, etc.).

4.3 Analysis of the results

A total of 8 algorithm were submitted to the first subtask, and our algorithm obtained the fourth place. Note that silence or no-chord segments were not estimated with our algorithm. The differences in the results between the participants are very small, probably because the approaches are similar (using HMM): Bello and Pickens obtained 66% of correct detected chords, Mehnert 65% correct, Ryynnen and Klapuri 64% correct, Papadopoulos and Peeters 63% correct, Khadkevich and Omologo 63% correct. The specificity of the approach we proposed is that there is no training at all whereas all the above-mentioned systems were pre-trained on the test set (Beatles songs). This is a very important point since it is very difficult and time-consuming to create labeled training data for audio chord detection. Our method was not among the fastest, however the running time is still interesting: about 20% of real time, which is a lot faster than most systems based on training. Our system compares also favorably to trained-system. Indeed, 7 algorithms were submitted to the second subtask. The approach proposed by Uchiyama, Miyamoto, and Sagayama gave results largely better than the other submitted algorithms (72% correct). Sheh and Ellis obtained 66% correct. All the remaining algorithms gave results above 62% correct.

5 REFERENCES

[1] H. Papadopoulos and G. Peeters. Large-scale study of chord estimation algorithms based on chroma representation and hmm. In *CBMI*, Bordeaux, France, 2007.

[2] H.Papadopoulos and G.Peeters. Simultaneous estmation of chord progression and downbeats from an audio file. In *ICASSP*, Las Vegas, 2008.

[3] T. Fujishima. Real-time chord recognition of musical sound: A system using common lisp music. In *ICMC*, pages 464–467, Beijing, China, 1999.

[4] G. Peeters. Musical key estimation of audio signal based on hmm modeling of chroma vectors. In *In DAFX, McGill*, pages 127–131, Montreal, Canada, 2006.

[5] J.C. Brown. Calculation of a constant q spectral transform. *J. Acoustical Society of America*, 89(1):425–434, 1991.

[6] G. Peeters. Chroma-based estimation of tonality from audio-signal analysis. In *ISMIR*, pages 115–120, Victoria, Canada 2006.

[7] Emilia Gómez. Tonal description of polyphonic audio for music content processing. *INFORMS J. on Computing*, 18(3):294–304, 2006.

[8] K. Noland and M. Sandler. Key estimation using a hidden markov model. In *ISMIR*, pages 121–126, Victoria, Canada, 2006.

[9] B. Gold and N. Morgan. *Speech and audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, Inc., 1999.