MIREX 2010: CHORD DETECTION USING A DYNAMIC BAYESIAN NETWORK

Matthias Mauch
National Institute of
Advanced Industrial Science and
Technology (AIST), Japan

Simon Dixon
Queen Mary University of London,
Centre for Digital Music,
London, UK

ABSTRACT

We present our submission *MD1* to the Chord Estimation Task of the 2010 Music Information Retrieval Evaluation eXchange (MIREX 2010). The front-end chroma generation is based on a new implementation of NNLS Chroma [1] as a Vamp plugin. The higher level model is implemented in Matlab as a dynamic Bayesian network (DBN) with extensive context modelling (metric position, key, chord and bass), leading to a very computationally complex method. The method has achieved the highest chord overlap ratio (0.8022) among the participating methods.

1. INTRODUCTION

We provide an outline of the NNLS (Non-Negative Least Squares) Chroma features in Section 2. Section 3 will explain the chord recognition methods. For results, please refer to the MIREX website [2] and to [5].

2. NNLS CHROMA

The chroma features are obtained using a prior approximate note transcription based on the non-negative least squares method (NNLS). For a more detailed description, see [5]. We first calculate a spectrogram (parameters, see Table 1), and then map it to a spectrogram with bins that are linearly spaced in log-frequency (similar to a constant-Q transform), with a resolution of three bins per semitone.

2.1 Tuning

As is frequently done in chord- and key- estimation (e.g. [2]), we adjust this spectrogram to compensate for differences in the tuning pitch. The tuning is estimated from the relative magnitude of the three bin classes. Using this estimate, the log-frequency spectrogram is updated by linear interpolation to ensure that the centre bin of every note corresponds to the fundamental frequency of that note in equal temperament.

[1] http://www.isophonics.net/nnls-chroma
[2] http://nema.lis.illinois.edu/nema_out/mirex2010/results/ace/

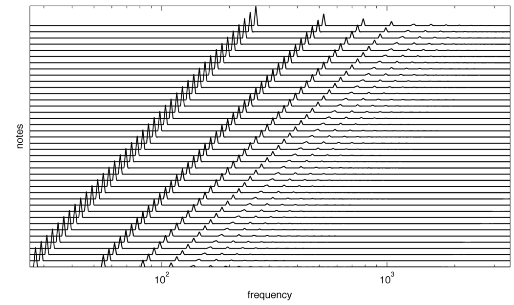

Figure 1: The first fourty elements of the chord dictionary.

2.2 Standardization

The spectrogram is then updated again to attenuate broadband noise and timbre by first subtracting a smoothed spectrum (negative values are set to 0) and dividing by an estimate of the local (in frequency) standard deviation of the amplitude.

2.3 NNLS

To determine note activation values we assume a linear generative model in which every frame Y of the log-frequency spectrogram can be expressed approximately as the linear combination $Y \approx Ex$ of note profiles in the columns of a dictionary matrix E, multiplied by the activation vector x. Finding the note activation vector that approximates Y best in the least-squares sense subject to $x \geq 0$ is called the non-negative least squares problem (NNLS). We choose a semitone-spaced note dictionary with exponentially declining partials, and use the NNLS

sampling rate	11025 Hz
frame length	4096 samples (≈372ms)
window	Hamming
step size	512 samples (≈46ms)

Table 1: Signal processing parameters.

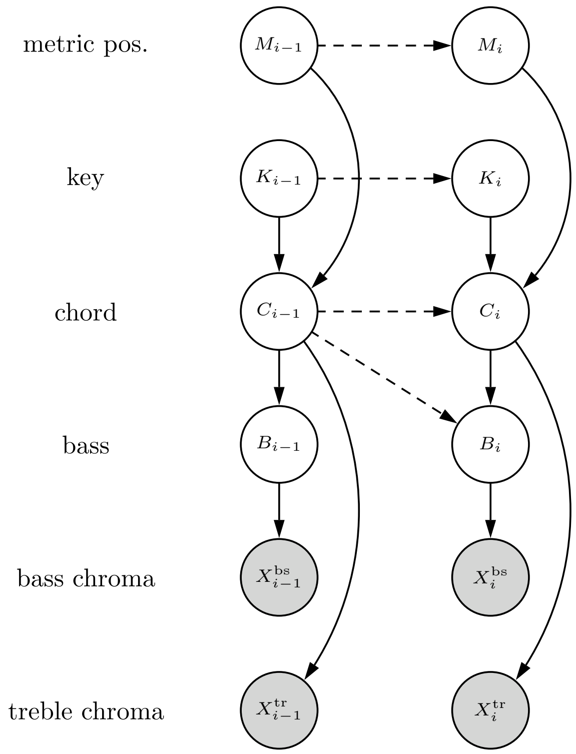

Figure 2: Our DBN model topology as a 2-TBN with two slices and six layers. Hidden nodes are clear, observed nodes are grey. Intra-slice dependency edges are drawn solid, inter-slice dependency edges are dashed.

algorithm proposed by Lawson and Hanson [3] to solve the problem and obtain a unique activation vector.

2.4 Treble and Bass Chroma Mapping.

In order to map the note activation to treble and bass chroma vectors we choose different profiles: the bass profile emphasises the low tone range, and the treble profile encompasses the whole note spectrum, with an emphasis on the mid range. The note activation vector (weighted by one of the profiles) is then mapped to the twelve pitch classes C,...,B by summing the values of the corresponding pitches.

3. CHORD TRANSCRIPTION USING A DYNAMIC BAYESIAN NETWORK

In order to transcribe the chords, we first generate a beat-synchronous chromagram from the NNLS Chroma output and then perform inference using the Viterbi algorithm on a dynamic Bayesian network, in which the beat-synchronous chroma are the observed variables.

3.1 Beat-Synchronization

In order to obtain beat times we use an existing automatic beat-tracking method [1]. A beat-synchronous chroma vector can then be calculated for each beat by taking the median (in the time direction) over all the chroma frames whose centres are situated between the same two consecutive beat times.

3.2 The Dynamic Bayesian Network

The two beat-synchronous chromagrams are now used as observations in the DBN, which is a graphical probabilistic model similar to a hierarchical hidden Markov model (originally proposed in [6]), see Figure 2. Our DBN jointly models metric position, key, chords and bass pitch class as hidden variables. The parameters are set manually according to musical expert knowledge, which is extensively described in [4]. We provide here a very brief description of the chord node, which is of central importance because its distribution is conditioned on metric position (influences chord change behaviour) and key (higher prior on chords belonging to the key), and the distributions of bass note node and treble chroma node are directly dependent on the chord note. The chord node models 121 chords of 11 different categories: major, minor, diminished, augmented, dominant 7th, minor 7th, major 7th, major 6th, and major chords in first and second inversion, and a "no chord" type.

The most likely sequence of hidden states is inferred using the beat-synchronous chromagrams of the whole song and the Viterbi algorithm.

4. ACKNOWLEDGEMENTS

This research was mainly funded by OMRAS2 (EPSRC grant EP/E017614/1). Thanks to Chris Cannam for helping with issues surrounding the Vamp implementation.

5. REFERENCES

[1] M. E. P. Davies, M. D. Plumbley, and D. Eck. Towards a musical beat emphasis function. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009)*, 2009.

[2] C. Harte and M. Sandler. Automatic chord identifcation using a quantised chromagram. In *Proceedings of 118th Convention*. Audio Engineering Society, 2005.

[3] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*, chapter 23. Prentice-Hall, 1974.

[4] M. Mauch. *Automatic Chord Transcription from Audio Using Computational Models of Musical Context*. PhD thesis, Queen Mary University of London, 2010.

[5] M. Mauch and S. Dixon. Approximate note transcription for the improved identification of difficult chords. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 135–140, 2010.

[6] M. Mauch and S. Dixon. Simultaneous estimation of chords and musical context from audio. *to appear in IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1280–1289, 2010.