MIREX AUDIO CHORD DETECTION

Maksim Khadkevich
FBK-Irst, Università degli studi di
Trento, Trento, Italy
khadkevich@fbk.eu

Maurizio Omologo
Fondazione Bruno Kessler - irst
Via Sommarive,18 - Povo - 38050
Trento, Italy omologo@itc.it

ABSTRACT

This paper describes the FBK-Irst system submitted to the MIREX'08 (Music Information Retrieval eXchange) Audio Chord Detection task. Because chords are major structural elements in tonal music, tools for automatic chord recognition can be widely used for indexing and retrieval in large audio databases, as well as for harmonic analyses of musical content. Our approach is based on a quite popular technique which uses chroma vectors as feature set and Hidden Markov Models (HMM) as statistical classifier. Feature extraction, training and testing methods, recognition results are briefly described.

1. FEATURE SET

In our analysis, the signal is down-sampled to 11025Hz, and converted to the frequency domain by a DFT using a Hamming window of length 0.37s with 50% overlap. We consider the range of frequencies between 100 Hz and 2 kHz. We use 12-dimensional chroma vectors [1] as acoustic features, which represent the intensities of the 12 semitone pitch classes. Each element of a vector corresponds to one of the 12 pitch classes. It is calculated as the sum of power at frequencies of its pitches over all octaves.

2. HIDDEN MARKOV MODELS

2.1. Training

For training Hidden Markov Models (HMM) chroma vectors are used as features. As opposed to many existing approaches ([3], [4], [6]), where chord is represented as a hidden state in one ergodic HMM, a separate model is created for each chord. Observation vector probabilities in each state can be characterized by a number of Gaussians in 12 dimensions, described by its mean vector and covariance matrix. It is assumed that components of feature vectors are uncorrelated with each other, so the covariance matrix has diagonal form. For each observation we use 512 12-dimensional Gaussian mixtures. Further increase from 512 to 1024 or 2048 mixtures leads to worse recognition rate, since the system learns training dataset and can not generalize pretty well. Figure 1 depicts chord recognition rate as a function of the number of Gaussians.

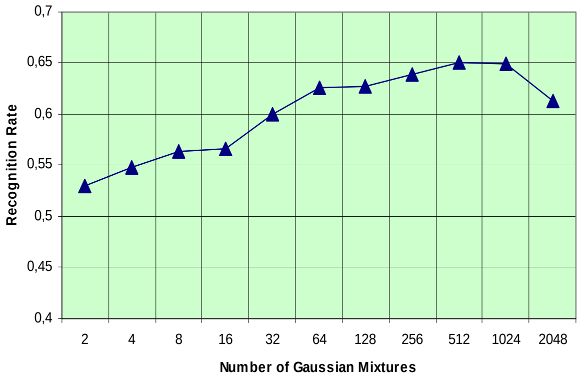

Figure 1. Recognition rate as a function of the number of Gaussians

Our approach has much in common with the approach of Sheh and Ellis [4]. They tried to recognize 147 different chords, including augmented and diminished. We reduce this number to 24 chords (major and minor for each of 12 roots). In order to prevent from lack of training data (some chord types can appear only few times in training corpus) only 2 models are trained: C-major and C-minor. For this purpose all chroma vectors obtained from labeled segments are mapped to the C-root using circular permutation. Then mean vectors and covariance matrixes are estimated for the 2 models. All the other models can be obtained by circular permutation procedure. Training is performed using the specific application of EM the expectation maximization (EM) algorithm – the Baum-Welch, or forward-backward algorithm. Ground-truth database annotations of 12 Beatles albums made by C. A. Harte [5] were used for training and testing purposes.

2.2. Recognition

Before running the recognition task, feature vectors are extracted from test data. There is no preliminary segmentation like in the train data, where chroma vector sequence was extracted for each chord segment. Only one sequence of chroma vectors is obtained for the whole test song. The trained chord models are connected as shown in Figure 2.

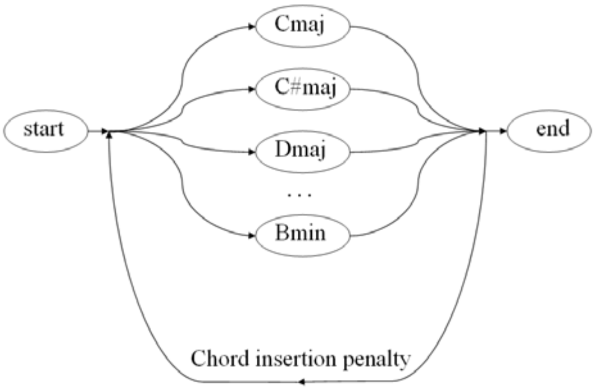

Figure 1. Connection scheme of trained models for decoding

The obtained connected model is used to determine a chord labeling for each song. The Viterbi algorithm [2] is used to recognize chords from test data. Varying the chord insertion penalty allows one to obtain output labels with different average segmentation lengths and different recognition rates. Figure 3 depicts chord recognition rate as a function of the chord insertion penalty.

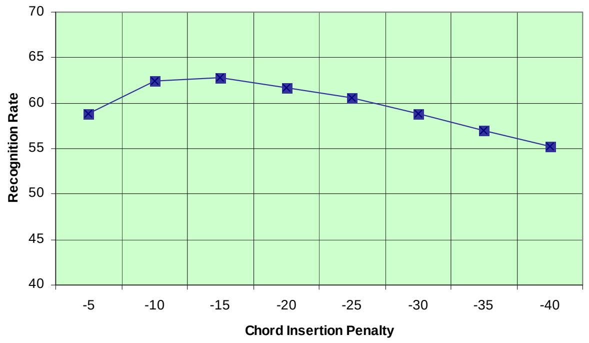

Figure 3. Recognition rate as a function of the chord insertion penalty

3. IMPLEMENTATION AND EXPERIMENTS

Feature extraction system was implemented in Java Programming Language. The Hidden Markov Model Toolkit (http://htk.eng.cam.ac.uk/) was used to perform training and testing. In the audio chord detection with pretrained systems subtask large collection of audio data was used to train models. In the n-fold test/train subtask the training was performed on a set of approximately 120 Beatles songs. Evaluation showed 63% in the first subtask and 55% in the second subtask. As for the future work, the system performance might be improved involving beat structure knowledge, bass and key information.

4. ACKNOWLEDGEMENTS

Many thanks to C. Harte for providing labeling data of the Beatles songs and to IMIRSEL for the great deal of effort spent in running the evaluation.

5. REFERENCES

[1] Goto, M. "A Chorus-Section Detecting Method for Musical Audio Signals", *Proc. ICASSP*, V, pp.437–440, 2003.

[2] L. R. Rabiner "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE.* 77, 257-286.

[3] 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.

[4] 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.

[5] C. A. Harte, et al., "Symbolic Representation of Musical Chords: A Proposed Syntax for Text Annotations," *Proc. ISMIR*, pp. 66-71, 2005.

[6] K. Lee and M. Slaney, "Automatic chord recognition using an HMM with supervised learning," in *Proceedings of the International Conference on Music Information Retrieval*, Victoria, Canada, 2006.