MIREX AUDIO CHORD DETECTION

Xinglin Zhang

Dept. of Computer Science
University of Regina
SK, Canada
zhang46x@cs.uregina.ca

Colan Lash

Dept. of Computer Science
University of Regina
SK, Canada
lash111c@uregina.ca

ABSTRACT

This extended abstract briefly introduces the submission to the 2008 Music Information Retrieval Evaluation eXchange in the Audio Chord Detection task. The chord vocabulary we are interested in includes only major and minor triads and 'N' for non-chord. To represent a chord for a frame, a traditional 12-dimension PCP vector is used and Hidden Markov Models are used as the classifier. The probability of the distribution of the PCP vector is modeled by a 12-dimension multivariate Gaussian. Also, the parameters of the Gaussian and the state transition matrix are learned through supervised learning. Tests were conducted on Beatles music and the result is shown.

Keywords: MIREX, Audio Chord Detection.

1. FEATURE VECTOR

The 12-dimension PCP vector, or the chroma vector, is used to represent a chord. At the first stage, the FFT is used to transform the signal from the time domain into the frequency domain, and then the information in the frequency domain is grouped by pitch class to get the PCP vector. Taking the FFT resolution into concern and in order to avoid the negative effects of the percussions and high harmonics of the notes, we restrict the frequency range to be from 65.41Hz(C2) to 1568Hz(G6).

2. SEGMENTATION

Based on the fact that the shortest chord length in all the Beatles music is around 0.2 second, we use a frame length of 0.2 second and an overlapping rate of 50% for both the training samples and the testing samples. In training, the WAV files are segmented as described above and the chord label for each frame are assigned according to the labeled file. In the situation where a frame happens to include the boundary of two different chords, the chord that has the larger portion in that frame is chosen as the chord.

3. HIDDEN MARKOV MODELS

3.1. Format of Parameters

Hidden Markov Models with Gaussian Mixture Models are used as the classifier. Based on the assumption that the distribution of the feature vector could be represented by a single Gaussian, we set the number of Gaussian Mixtures to be 1. Furthermore, we make an assumption that the covariance matrix for each state is diagonal, which means the 12 dimensions are independent to each other, and the value of the diagonal elements are simply the variance of each dimension.

3.2. Supervised Learning

All the parameters of the HMMs are estimated through supervised learning. The mean vector for each state (chord) is obtained by calculating the mean of all the samples observed for that state; the covariance matrix for each state is obtained by calculating the covariance matrix of all the samples observed for that state. A full covariance matrix is used.

The state transition probability is obtained by first counting the numbers of chord transitions for each chord and then normalizing.

4. CHORD SEQUENCE

Since the HMMs do not contain information of "N" chords, a pre-processing which detects silence is taken first to find the non-chord frames. The chord sequence then consists of : non-chords obtained by pre-processing and major and minor triads obtained by the Viterbi algorithm.

5. RESULT

5.1. Training

We use 10 albums of Beatles music for training. The corpus contains 140 songs. Table 1 shows the details of training.

Artist	The Beatles
Number of albums	10
Number of songs in total	140
Platform	CPU: Xeon 3.20Ghz RAM: 2.0G OS: Windows XP
Time(feature extraction and parameter estimation)	20.3 mins

Table 1. Training Details

5.2. Testing

One album is used to test the trained system. We do not do any onset detection, thus the output is based on a per-frame basis. As well, the testing is conducted on the same machine used for training. Table 2 shows the time used and the frame correction rate (FCR).

Name of Song	Time used	FCR
Magical_Mystery_Tour	11.57s	63.4%
The_Fool_On_The_Hill	11.92s	30.0%
Flying	8.63s	29.8%
Blue_Jay_Way	15.34s	71.7%
Your_Mother_Should_Know	9.84s	55.2%
I_Am_The_Walrus	18.43s	48.0%
Hello_Goodbye	13.08s	54.91%
Strawberry_Fields_Forever	16.07s	31.58%
Penny_Lane	11.71s	47.98%
Baby_You're_A_Rich_Man	11.76s	60.17%
All_You_Need_Is_Love	15.25s	32.14%

Table 2. Testing results

6. REFERENCES

[1] M. Goto. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*, 30(2):159–171, 2001.

[2] A. Sheh and D. P. Ellis. Chord segmentation and recognition using em-trained hidden markov models. In *Proceedings of the International Symposium on Music Information Retrieval*, Baltimore, MD, 2003.

[3] K. Lee and M. Slaney. A unified system for chord transcription and key extraction using hidden markov models. In *Proceedings of International Conference on Music Information Retrieval*, 2007.