AUDIO GENRE CLASSIFICATION USING RHYTHM AND BASS-LINE PATTERN INFORMATION

Emiru Tsunoo† **Geroge Tzanetakis**†† **Nobutaka Ono**† **Shigeki Sagayama**†

† Graduate School of Information Science and Technology University of Tokyo, Japan

{tsunoo, onono, sagayama}@hil.t.u-tokyo.ac.jp

†† Computer Science Department University of Victoria, Canada

gtzan@cs.uvic.ca

ABSTRACT

This paper discusses an approach for the feature extraction for audio genre classification and many other tasks of music information retrieval (MIR). Many musical genres are characterized not only by timbral information but also by temporal features such as rhythm patterns and bass-line patterns. In particular, modern music pieces mostly have certain fixed rhythm and bass-line patterns per genre, and the extraction of these representative patterns is thought to be useful for genre classification. We have proposed the extraction of rhythm and bass-line patterns, and using this method, a new feature extraction system was implemented into Marsyas which is a open source software framework for audio analysis, synthesis and retrieval for MIR.

1. INTRODUCTION

Due to the increasing size of music collections available on computers and portable music players, researches related to music information retrieval (MIR) including automatic genre classification from audio. Genre classification has been provided a structured way of evaluating new representations of musical content through a long time in MIREX. In this task, not only instrumental information but also rhythm and bass-line informations are thought to be important. For instance, talking about rhythm pattern, samba and tango have very similar timbre features but totally different rhythmic patterns. About bass-line patterns, bass parts in most rock songs consist of root notes of the chords in a uniform rhythm, and comparison to this, bass parts in jazz songs have a lot of characteristic movements which are called walking bass. If such representative bar-long rhythm and bass-line patterns in music can be captured automatically as templates, they can potentially be used to characterize different music genres directly from audio signals.

In previous research, timbral features, rhythmic features and pitch features have been used for audio genre classification [1]. In this work, the timbral features were the most dominant and the pitch features used were not limited to the bass register. Research more closely related to extracting rhythmic pattern are work with rhythmic patterns includes Dixon [2] which extract a periodical pattern from acoustic signals heuristically and Peeters [3] which extracts features based on the periodicity of the spectrum. These approaches can successfully discriminate styles such samba or tango primarily based on rhythmic information. Studies more related to bass part extraction are represented by Goto [4] who modeled musical melodic and bass notes with Gaussian mixture model (GMM) and estimated both of them. Using this pitch estimation method, Marolt [5] worked on the clustering of melodic lines using GMM. Researches using bass-line information for genre classification include [6] and [7]. However the bass-line features discussed here were based on overall statistics and did not represent directly temporal information.

In this paper, we discuss an approach for extracting unit rhythm and bass-line patterns from a number of audio tracks and propose a feature vector for the application to genre classification. First, we separate percussive sounds and harmonic sounds of the audio tracks and estimate bar lines which divide tracks into measures. Then we propose a clustering method specialized to rhythm patterns using a combination of dynamic programming and k-means clustering, and to bass-line patterns based on the k-means clustering algorithm. For the purpose of an application to audio genre classification, the scheme to extract feature vector based on extracted patterns which contain temporal information is suggested. All process is implemented using an open source software *Marsyas*[1] which is open source software with specific emphasis on Music Information Retrieval (MIR) [8].

2. HARMONIC/PERCUSSIVE SOUNDS SEPARATION

Generally, harmonic and percussive sounds are mixed in the observed spectrograms of audio pieces. Therefore in order to perform rhythm and bass-line pattern analysis it is useful to separate these components as a preprocessing, percussive components for rhythm analysis and harmonic components for bass-line analysis. We utilize the harmonic/percussive sound separation (HPSS) technique proposed by Ono [9] that is based on the difference of general timbral features. By looking at the upper left figure in Fig. 1, a typical instance of spectrogram, one can observe that harmonic components tend to be continuous along the temporal axis in particular frequencies. On the other hand, percussive components tend to be continuous along the frequency axis and temporally short. Mask functions for separating the two components (harmonic and percussive) are calculated following a maximum a priori (MAP) estimation approach using the expectation maximization (EM) algorithm. Applying this approach to the shown spectrogram, harmonic and percussive components are separated

[1] http://marsyas.sness.net/

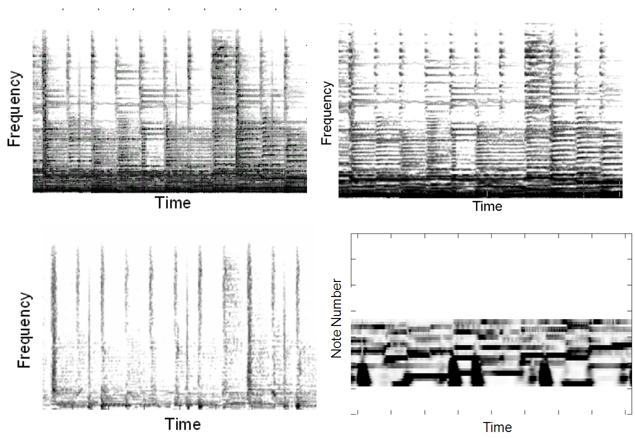

Figure 1. The original spectrogram (upper left), the harmonics-emphasized spectrogram (upper right) and the percussion-emphasized spectrogram (lower left) of a popular music piece (RWC-MDB-G-2001 No.6 [10]). The low-pass filtered logarithmic spectrogram calculated using wavelet transform from harmonics-emphasized spectrogram is shown in lower right figure.

(harmonic and percussive components are shown in the upper right and the lower left of Fig. 1 respectively).

3. BAR-LONG UNIT PATTERN CLUSTERING

3.1 Rhythm Pattern Clustering

Bar-long percussive patterns are frequently common and characteristic of a particular genre or style. Automatically detecting these patterns is a "chicken-and-egg" problem in that sets of bar-long unit rhythm patterns may be determined only after their corresponding unit boundaries in the music pieces are given, and vice versa. This is complicated by tempo fluctuations which might cause the unit pattern to stretch or shrink.

In order to solve these problems, the rhythm map which we have proposed [11, 12] is employed. The rhythm map is an approach to estimate representative bar-long percussive patterns and its segmentation (i.e. bar lines) simultaneously. This algorithm also requires the source separation method shown in previous subsection as a preprocessing and deals with the percussive-emphasized spectrogram. By iterating dynamic programming (DP) matching and updating the templates used for DP matching based on the k-means-clustering-like update rules, both segmentation and templates themselves are updated. After convergence, the multiple percussive patterns in the input song are learned and the optimal segmentation is obtained. We use this estimated segmentation as bar lines. Fig. 3 illustrates the flow of this algorithm.

3.2 Bass-line Pattern Clustering

Bar-long bass-line patterns are also frequently common and characteristic of a particular genre. Bar lines have been already estimated in the process shown in section 3.1. Therefore, only problem to be solved is that unit bass-line patterns are shifted in pitch according to the chord played. For example, a pattern consists of only root notes in a uniform rhythm, all notes in this pattern need to be pitch-shifted by the same amount of notes according to the chord, because the root note changes accompany with the chord changes. Fig. 3 illustrates the flow of the algorithm.

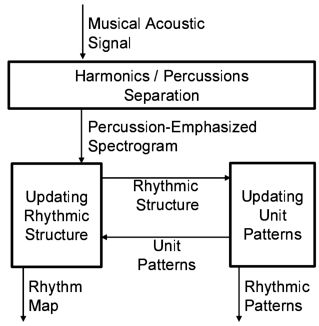

Figure 2. The flow diagram of the extraction of rhythm patterns.

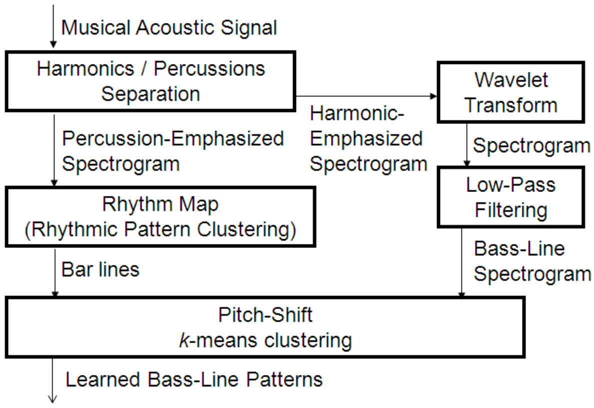

Figure 3. The flow diagram of the bass-line pattern extraction.

If there was no pitch shift (problem III) a simple k-means clustering approach can be used to estimate the representative bar-long bass-line patterns: Distances between each bar-long spectrogram pattern and centroid spectrogram patterns are calculated, and the centroids are updated by averaging the sample patterns. In order to deal with pitch-shift problem, we have proposed an approach where every possible pitch-shift is compared in k-means framework [13]. The lower right of Fig. 1 shows the logarithmic spectrogram which is processed low-pass filtering after Gabor wavelet transform whose frequency resolution is semitone (100 cents). The low-pass filtering (actually a band-pass filtering) was done by setting high and low frequency components to be zero. This kind of spectrogram is used for this purpose.

4. FEATURE EXTRACTION

4.1 Genre Classification via Pattern Clustering

Ideally percussive patterns for a particular genre or style would be fixed and would be automatically extracted perfectly. If that was the case then automatic genre classification could be performed simply by looking at which particular rhythm and bass-line patterns are used in a music piece. However in practice there is no guarantee that patterns are fixed for a particular genre/style or that their automatic extraction will be perfect. Therefore in many cases the percussive and bass-line patterns of a particular music piece will belong to more than one genre. To address these problems simultaneously we utilize a pattern occurrence histogram and distance related representation followed statistical machine learning to automatically classify music genre. Supervised learning classifiers such as Support Vector Machines (SVM) [14] can be used for this purpose.

4.2 Percussive Pattern Feature Extraction

One possible way to extract feature vector is count up which percussive pattern templates are contained in a song and calculating the genre pattern occurrence histogram, similarly to Latent Semantic Indexing approach [15].

If M pattern templates are learned from genre g ($g = 1, \ldots, G$), an alignment can be calculated using dynamic programming to get the templates $T_{m,g}$ that exist in the song s. Then, the occurrence number of the patterns from genre g can be simply calculated by summation as follows:

$$c_{s,g} = \sum_{m=1}^{M} c_{s,m,g} \tag{1}$$

where $c_{s,m,g}$ is the number of the template $T_{m,g}$ in the song s, and the eventual G dimensional pattern occurrence histogram features vector $\boldsymbol{x}$ of song s can be written as

$$x_g = \frac{c_{s,g}}{N_s} \tag{2}$$

which is normalized by N_s, the number of measure in the song s.

4.3 Bass-line Pattern Feature Extraction

For bass-line pattern, one way to extract feature vector is calculating distances between every measure of input spectrogram and every template pattern and averaging them through whole an input piece. Even though there is a possibility for a music piece to belong to more than one genre templates, the distances between spectrogram in the input piece and learned templates are still affected, e.g., one measure spectrogram in blues song is close enough to the templates learned from blues collection even if its distance is not the smallest.

The mathematical definition of the feature vector is following. After bass-line pattern templates are learned, we have $K \cdot G$ templates when K templates are learned from G genres. Then the distances between input song which have M measures and learned templates are calculated. The averaged distances are obtained as follows:

$$d_l = \frac{1}{M} \sum_{m=1}^{M} D(X_m, B_l) \tag{3}$$

where $1 \leq l \leq KG$ is the template number. The feature vector $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = (d_1, d_2, \ldots, d_{KG})^T. \tag{4}$$

We use this feature vector for a supervised learning classification to classify music genre.

5. CONCLUSIONS

We discussed an approach for clustering bar-long common percussive patterns and bass-line patterns for particular genres/styles and extracting feature vectors for genre classification. We used HPSS technique to separate percussive components and harmonic components from audio signals as a preprocessing. Percussive patterns were clustered using a combination of one-pass DP and k-means clustering algorithm. Bass-line patterns were clustered using a new clustering method based on k-means clustering with pitch-shift. For audio genre classification, new feature vectors were defined as pattern occurrence histograms for percussive patterns and as averaged distances from each template for bass-line patterns.

Future work includes using n-gram model approach rather than only looking at the uni-gram histogram. Additionally, other features than pattern distance vector can be devised. Combination with other features like chord transition information can be done to improve further genre classification as well.

6. REFERENCES

[1] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transaction on Speech and Audio Processing*, 10(5):293–302, 2002.

[2] S. Dixon, F. Guyon, and G. Widmer. Towards characterization of music via rhythmic patterns. In *Proc. of the 5th Int. Conf. on Music Information Retrieval*, pages 509–516, 2004.

[3] G. Peeters. Rhythm classification using spectral rhythm patterns. In *Proc. of the 6th Int. Conf. on Music Information Retrieval*, pages 644–647, September 2005.

[4] M. Goto. A real-time music-scene-description system: Predominant-f0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43(4):311–329, 2004.

[5] M. Marolt. Gaussian mixture models for extraction of melodic lines from audio recordings. In *Proc. of ISMIR*, pages 80–83, 2004.

[6] C. McKay and I. Fujinaga. Automatic genre classification using large high level musical feature sets. In *Proc. of ISMIR*, pages 525–530, 2004.

[7] Y. Tsuchihashi, T. Kitahara, and H.Katayose. Using bass-line features for content-based mir. In *Proc. of ISMIR*, pages 620–625, 2008.

[8] G. Tzanetakis. *Marsyas-0.2: A Case Study in Implementing Music Information Retrieval System*, chapter 2, pages 31 – 49. Idea Group Reference, 2007. Shen, Shepherd, Cui, Liu (eds).

[9] N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama. A real-time equalizer of harmonic and percussive componets in music signals. In *Proc. of the 9th Int. Conf. on Music Information Retrieval*, pages 139–144, September 2008.

[10] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. Rwc music database: Music genre database and musical instrument sound database. In *Proc. of the 4th Int. Conf. on Music Information Retrieval*, pages 229–230, October 2003.

[11] E. Tsunoo, N. Ono, and S. Sagayama. Rhythm map: Extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals. In *Proc. of ICASSP*, pages 185–188, 2009.

[12] Emiru Tsunoo, George Tzanetakis, Nobutaka Ono, and Shigeki Sagayama. Audio genre classification using percussive pattern clustering combined with timbral features. In *Proc. of ICME*, pages 382 – 385, 2009.

[13] Emiru Tsunoo, Nobutaka Ono, and Shigeki Sagayama. Musical bass-line pattern clustering and its application to audio genre classification. In *Proc. of ISMIR*, 2009.

[14] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[15] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.