RHYTHM MAP: EXTRACTION OF UNIT RHYTHMIC PATTERNS AND ANALYSIS OF RHYTHMIC STRUCTURE FROM MUSIC ACOUSTIC SIGNALS

Emiru Tsunoo, Nobutaka Ono and Shigeki Sagayama

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyou-ku, Tokyo, 113-8656, Japan
{tsunoo,onono,sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

This paper discusses an approach to extract constituent percussive bar-long patterns in a music piece given as acoustic signal and to analyze the music structure with a map of constituent rhythmic patterns. Possible applications include music genre classification, music information retrieval (MIR) and music modification such as replacing rhythmic patterns with others. We propose a mathematical method based on One-pass DP algorithm and k-means clustering to extract unit percussive rhythmic patterns. As the result of identifying and localization the unit patterns in the entire piece, we obtained a music structure in the form of a map of rhythmic patterns.

Index Terms— Percussive sound, Spectral analysis, Dynamic programming, Pattern clustering method, feature extraction

1. INTRODUCTION

For the purpose of music information retrieval, rhythm-related features have the high potential in characterization of music, particularly related to genres. For example, the distinction between samba and tango exists primarily in their bar-long rhythmic patterns.

Obviously, rhythm is one of the most fundamental elements of music to characterize it. From a microscopic viewpoint, unit rhythmic patterns are often the elements to form measures and beats. From a macroscopic viewpoint, multiple rhythmic patterns included in a whole music piece very often form the entire structure of the piece. If possibly multiple unit rhythmic patterns in the music piece can be extracted properly and the music structure can be analyzed in terms of the use of unit rhythmic patterns, they will be helpful in characterization of the music and then be useful in music genre identification and music information retrieval.

In the past research, the most fundamental aspect of rhythm analysis research was related to beat tracking [1] . Another related work was extracting beat histogram which was a rhythmic content feature, and discussed precisely in [2]. Another work related to rhythmic patterns dealt with measuring the similarity of rhythmic patterns [3]. In this work, spectral features were extracted and these patterns were compared using dynamic time warping (DTW). However, it was not successful in application to the real music pieces. Works which dealt with extracting rhythmic patterns are typified by [4] which extracts a periodical pattern from acoustic signals heuristically, and by [5] which extracts features based on the periodicity of spectrum. They successfully discriminate genres between the rhythmical songs like samba and tango.

2. RHYTHM SEGMENTATION AND LABELING

2.1. Four Fundamental Problems in Rhythm Analysis

The problem here is a "chicken-and-egg" problem: a set of fundamental bar-long rhythmic patterns may be determined only after unit boundaries in the music piece are given, and vice versa, i.e., unit boundaries can be determined only after unit patterns are given. Another problem is that the tempo may fluctuate. The unit rhythm pattern may stretch or shrink. There is another practical problem that is typical in music (especially, modern popular music and jazz) - it contains harmonic sounds, which can disturb the spectrogram-based rhythm analysis.

Therefore the problems in extraction of rhythmic unit patterns from the input music signal can be summarized as consisting of the following four problems:

(i) the input acoustic signal may contain not only percussive sounds but also melodic/chordal sounds,
(ii) there may be fluctuations in tempo and in pattern itself made by the performer,
(iii) unit segmentation is unknown, and
(iv) unit rhythmic patterns are unknown.

In the next a few subsections, we discuss an approach to solve these four problems.

2.2. Emphasizing Percussive Components

The first problem is that harmonic and percussive sounds are mixed in the observed spectrogram. To achieve separation of these two components without prior knowledge, we can employ Ono's method [6] to decompose music signal into har-

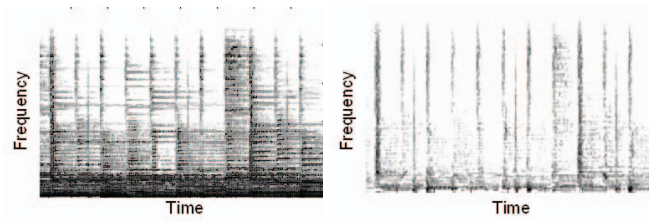

Fig. 1. The original spectrogram (left) and the percussion-emphasized spectrogram (right) of a popular music piece (RWC-MDB-G-2001 No.6).

monic and percussive components which would emphasize percussive rhythmic patterns.

The left side of Fig. 1 shows a typical instance of spectrogram. Generally, harmonic components in the spectrogram tend to be continuous along the time axis near particular frequencies, i.e., fundamental frequencies and their overtones. On the other hand, percussive components tend to be continuous along the frequency axis while temporally short. Having performed the EM algorithm to estimate mask functions, we can then use them to separate percussive sounds from the audio input signal. By applying this algorithm to the spectrogram on the left side of Fig. 1, we were able to separate the harmonic and percussive components, and the emphasized percussive components are shown on the right side of Fig. 1.

2.3. Iterative Update of Rhythmic Structure and Unit Patterns

If the true set of unit rhythmic patterns is given as templates, the problem (iii) is parallel to the continuous speech recognition problem where One-Pass DP (Dynamic Programming) algorithm [7] can be employed to find the sequence of uttered words. Accordingly, One-Pass DP divides a music piece into segments, each optimally corresponding to template patterns. Also, because of its flexibility in time alignment, the problem (ii) is solved simultaneously.

The problem (iv)—the need to estimate each of fundamental unit rhythm pattern—is a chicken-and-egg problem if both segmentation and unit rhythmic patterns are unknown. For that reason, it is necessary to estimate segmentation and unit patterns simultaneously. While this kind of unsupervised training problems have been solved in various ways, here the k-means clustering algorithm in combination with the One-Pass DP algorithm is employed. Unit rhythmic patterns and the music structure are trained iteratively. Fig. 2 illustrates the flow of this algorithm. Considering a probabilistic model in which the input patterns are drawn from the template patterns according to a certain distribution, the problem is solved by maximizing the likelihood.

2.4. Rhythmic Structure Analysis by One-pass DP Algorithm

One-Pass DP algorithm gives the optimal segmentation of input spectral patterns by giving template patterns like in continuous speech recognition. As most of music pieces with percussive instruments keep their tempos, we can design the locally allowed paths in DP as in Fig. 3, to reduce the range of a time fluctuations.

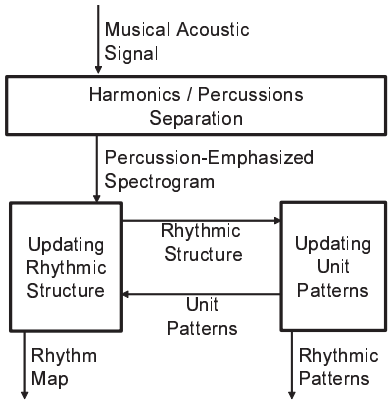

Fig. 2. The flow diagram of the system.

Assuming that the series of spectra is observed composing probabilistic distributions which expect corresponding spectra in templates, a probabilistic model can be designed, and we can assume that the sound is close to that in templates when its output probability is high.

Percussive spectral time-frequency components are defined as $P(t_x, f_n) = P(x, n)$ where t_x is time and f_n is logarithmic frequency. Spectral components at a time t_x are defined as a vector $\boldsymbol{r}_x = (P_{x,1}, \ldots, P_{x,N})^T$. Similarly, template spectrogram composed of components $M_m(t_i, f_n) = M_{m,i,n}$ $(m = 1, \ldots, M)$ are defined as series of vectors $\boldsymbol{\mu}_{m,i} = (M_{m,i,1}, \ldots, M_{m,i,N})^T$.

The output probability of the vector $\boldsymbol{r}_x$ from the spectrum of the frame i in template m is written as

$$p_{m,i}(\boldsymbol{r}_x) = \frac{1}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}_{m,i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{e}_{m,i,x}^T \boldsymbol{\Sigma}_{m,i}^{-1} \boldsymbol{e}_{m,i,x}\right) \quad (1)$$

where $\boldsymbol{e}_{m,i,x} = (\boldsymbol{\mu}_{m,i} - \boldsymbol{r}_x)$, and $\boldsymbol{\Sigma}_{m,i}$ is the diagonal covariance matrix for the time i in the template m.

According to One-Pass DP algorithm, logarithmic likelihoods $\ln(p_{m,i}(\boldsymbol{r}_x))$ are multiplied by the weight w and summed up one after another. The weight can be designed as depicted in Fig. 3. In the end, an alignment according to given templates is calculated by finding the likeliest path, and a music structure composed of rhythmic pattern templates is estimated.

The alignment calculated above gives a correspondence between the spectrum $r_{x(a)}$ of the time index $x(a)$ and the template spectrum $\boldsymbol{\mu}_{m(a),i(a)}$ of the time index $i = i(a)$ in template $m = m(a)$. Therefore, the summation of logarithmic likelihood can be written as

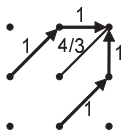

Fig. 3. Local continuity constraints with slope weighting.

$$D_A = -\frac{1}{2}\Big(\sum_{a=1}^{A}\big(N\log(2\pi)+\log|\mathbf{\Sigma}_{m(a),i(a)}|\big)\cdot w(a) + \sum_{a=1}^{A}\boldsymbol{e}_{m,i,x}(a)^T\mathbf{\Sigma}^{-1}_{m(a),i(a)}\boldsymbol{e}_{m,i,x}(a)\cdot w(a)\Big) \quad (2)$$

where $\boldsymbol{e}_{m,i,x}(a) = (\boldsymbol{\mu}_{m(a),i(a)} - \boldsymbol{r}_{x(a)})$.

2.5. Updating Unit Patterns by k-means Clustering

Next, like in k-means clustering, central patterns of each cluster are calculated and are set as new template patterns. Each template pattern is calculated by averaging segments labeled as the same cluster, keeping the alignment given by One-Pass DP algorithm. The total likelihood D_A calculated in One-Pass DP algorithm (Eq. 2) can be maximized by calculating such a weighted average. Therefore, the total likelihood D_A is increased in each update of the template patterns, and the convergence is guaranteed.

The template patterns are updated based on the maximum likelihood estimation and their parameters will be $\hat{\theta} = (\hat{\boldsymbol{\mu}}_{m,1},\ldots,\hat{\boldsymbol{\mu}}_{m,I_m},\hat{\mathbf{\Sigma}}_{m,1},\cdots,\hat{\mathbf{\Sigma}}_{m,I_m})$. Setting $\frac{\partial D_A}{\partial \mu_{m,i}}$ to 0, $\hat{\mu}_{m,i}$ is written as

$$\hat{\mu}_{m,i} = \frac{\sum_{a\in A_{m,i}}\boldsymbol{r}_{x(a)}\cdot w(a)}{\sum_{a\in A_{m,i}} w(a)} \quad (3)$$

where $A_{m,i} = \{a|m(a)=m, i(a)=i\}$. In a same way, setting $\frac{\partial D_A}{\partial \mathbf{\Sigma}_{m,i}}$ to 0, $\hat{\mathbf{\Sigma}}_{m,i}$ is written as

$$\hat{\mathbf{\Sigma}}_{m,i} = \frac{\sum_{a\in A_{m,i}}\boldsymbol{e}_{m,i,x}(a)\boldsymbol{e}_{m,i,x}(a)^T\cdot w(a)}{\sum_{a\in A_{m,i}} w(a)}. \quad (4)$$

Therefore, the total likelihood calculated after this update, D'_A, satisfies

$$D'_A \geq \hat{D}_A = \max_{\theta} D_A \geq D_A \quad (5)$$

and this iterative update never reduces the total likelihood, so the convergence is guaranteed.

2.6. Procedural Summary of the Algorithm

The discussed algorithms above are summarized in the following procedure:

1. Emphasis of Percussive Sounds: by using Ono's method, percussive components of the series of spectrogram are separated.

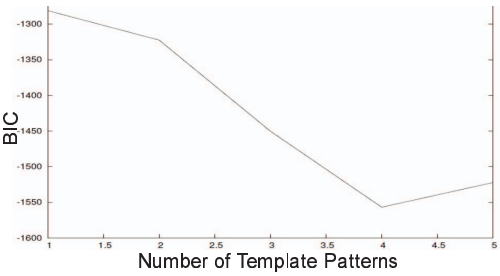

Fig. 4. The BIC calculated at every number of template patterns.

2. Giving the Initial Template Patterns for One-Pass DP Algorithm: the initial template patterns are made as average rhythmic patterns in an input music or as average patterns of typical rhythms in every genre.
3. Giving the Optimal Segmentation: using One-Pass DP algorithm, a segmentation is calculated from given reference patterns.
4. Update of the Reference Patterns: like in k-means algorithm, the centroids of the clusters are recalculated and updated as new reference patterns.
5. Iteration: repeat step 3 and 4 until the dissimilarity cost calculated in One-Pass DP algorithm converges.

3. EXPERIMENTAL EVALUATION

3.1. Data Set

The purpose of this experiment is to confirm that the algorithm above can be applied to real music pieces which contain harmonic components as well as percussive components.

We used WAV files from the RWC music database [8] down-sampled to 22.05 kHz single-channel files. In this process, we used a Hanning window with 1024 samples and 50% overlap at performing a Short Time Fourier Transform (STFT). Using Ono's method, we obtained the percussive spectral patterns and spectrum of each frame was summed up to 8 dimensional spectral vectors to reduce the computation time.

3.2. Results

We applied our algorithm to a dance music: RWC-MDB-G-2001 No. 16 in the data set above. We determined the number of template patterns by using the Bayesian information criterion (BIC). Fig 4 shows the calculated BIC at every number of templates, and it means the optimal number of templates is 4. Therefore we gave 4 initial template patterns and convergence of our algorithm, the alignment became the right side of Fig. 5 and the fundamental unit rhythmic patterns learned are illustrated on the left side of Fig. 5.

By listening to this music, we were able to tell that pattern 1 was repeatedly played and once in four measures, pattern 2 was played. Following such fundamental rhythms, an interval rhythmic pattern was played (pattern 3), followed by a pattern in the climax part (pattern 4). This can be clearly seen

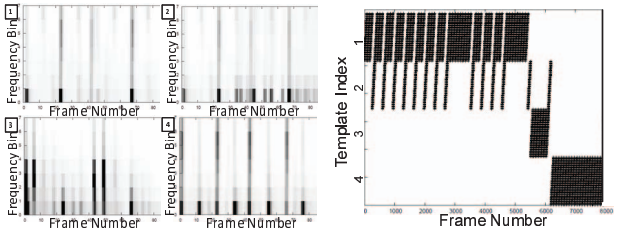

Fig. 5. 4 extracted percussive unit rhythmic pattern spectrograms from a dance music(No. 16) (left) and the corresponding alignment, i.e., "Rhythm Map" (right).

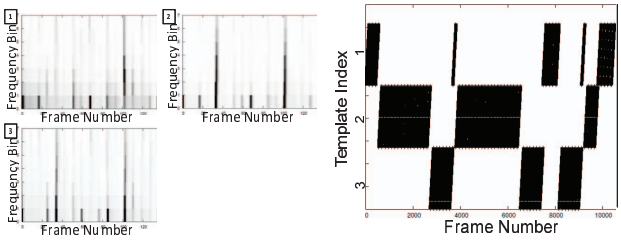

Fig. 6. 3 extracted unit rhythmic pattern spectrograms from a popular music(No. 6) (left) and the corresponding Rhythm Map (right).

on the right side of Fig. 5, which depicts a music structure in the form of a map of rhythmic patterns, which we named "Rhythm Map".

Another example of a popular music "Rhythm Map" (RWC-MDB-G2001 No. 6) is illustrated on the right side of Fig. 6, and the left side of Fig. 6 shows 3 corresponding rhythmic patterns.

3.3. Evaluation

We have evaluated how the proposed algorithm classifies the rhythmic patterns compared to how the humans do it. We applied the same algorithm to 4 music pieces: RWC-MDB-G-2001 No. 6, No. 16, No. 19, and No. 26.

We developed an interface and asked subjects to classify the rhythmic patterns with it to define "correct" answers. Then, we compared the rhythmic pattern labels we obtained and classified segmentation the algorithm estimated. For evaluation, we calculated the ratio of correctly classified frames.

The result is shown in Table 1. Even though this algorithm has a bootstrap problem, and so its accuracy depends on the initial patterns, basically we could confirm the validity of classification, which was close to how humans do it. Our algorithm worked particularly well for dance music which have strong percussive components. The soul music (No. 26) above had complex percussive patterns and the main reason for low accuracy was the rotation of the patterns.

Table 1. Percentage of correctly classified segments.

Music	Correct Frames (%)
No.6 (Pop)	83.32
No.16 (House)	97.87
No.19 (Techno)	87.25
No.26 (Soul)	69.73

4. CONCLUSIONS

We discussed an approach to extract unit rhythmic patterns of percussive components in music signals and to analyze and display the music structure in a map form. We used Ono's method to extract percussive components from music signals and proposed an algorithm with which the unit rhythmic patterns and "Rhythm Map" are learned iteratively using a combination of One-Pass DP and k-means clustering algorithms. Experiments over music pieces with percussive parts from various genres confirmed that our algorithm could extract proper rhythmic patterns and a rhythm map to locate the rhythmic patterns in the music piece.

Future works include improvement of segmentation into unit patterns, i.e. rotation problem, using additional clues such as transition of melodic and/or chordal components. We also plan to apply the method to more recordings. After that, this algorithm can be applied to the genre classification in which the template patterns are learned from every genre and the proportions of those representative patterns are extracted as a features vector using dynamic programming.

5. REFERENCES

[1] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, June 2001.

[2] G. Tzanetakis, G. Essl, and P. Cook, "Audio analysis using the discrete wavelet transform," in *Proc. of WSES Int. Conf. on Acoustics and Music: Theory and Applications*, 2001.

[3] J. Paulus and A. Klapuri, "Measuring the similarity of rhythmic patterns," in *Proceedings of the 3rd International Conference on Musical Information Retrieval*, IRCAM Centre Pompidou, 2002, pp. 150–156.

[4] 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*, 2004, pp. 509–516.

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

[6] 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*, September 2008, pp. 139–144.

[7] H. Ney, "The use of a one-stage dynamic programming algorithm forword recognition," in *Int. Conf. on Acoust., Speech, Signal Processing*, 1984, pp. 263–271.

[8] 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*, October 2003, pp. 229–230.