

# Repeating Pattern Extraction Technique(REPET);A method for music/voice separation.

Wakchaure Amol Jalindar<sup>1</sup>, Mulajkar R.M.<sup>2</sup>, Dhede V.M.<sup>3</sup>, Kote S.V.<sup>4</sup>

<sup>1</sup> Student, M.E(Signal Processing), JCOE Kuran, Maharashtra, India

<sup>2</sup> ME Co-ordinator, Electronics & Telecommunication, JCOE Kuran, Maharashtra, India

<sup>3</sup> HOD, Electronics & Telecommunication, JCOE Kuran, Maharashtra, India

<sup>4</sup> Lecturer, Computer Engineering, Jaihind Polytechnic Kuran, Maharashtra, India

## ABSTRACT

*With the popularity of multimedia applications, a large amount of music data has been accumulated on the Internet. Automatic classification of music data becomes a critical technique for providing an efficient and effective retrieval of music data. In this paper, we propose a new approach for classifying music Data based on their contents. In this approach, we focus on monophonic music features represented as rhythmic and melodic sequences. Moreover, we use repeating patterns of music data to do music classification. For each pattern discovered from a group of music data, we employ a series of measurements to estimate its usefulness for classifying this group of music data. According to the patterns contained in a music piece, we determine which class it should be assigned to. We perform a series of experiments and the results show that our approach performs on average better than the approach based on the probability distribution of contextual information in music.*

**Keyword** - Music classification, Repeating patterns, Feature extraction.

## 1. Introduction

As the amount of music data increases, classification of music data has become an important issue. In [2][6], the machine learning techniques including naïve Bayesian, linear, and neural network are employed to build classifiers for music styles. As a result, they identify emotional classes of music styles such as lyrical and frantic. Chai and Vercoe [4] classify folk music into groups based on melody, where each group corresponds to the music of a particular country. They first build a hidden Markov model for each country based on training data. After that, a music piece can be classified by The probabilities Associated with The model. In this paper, we first find useful information for classification from the symbolic representations of music data. A similarity measure considering human perception of music is then designed to measure the similarity degree between two music objects. Finally, we consider a broader coverage of music with seven classes to do performance Evaluation. To represent the music data, a variety of symbolic features, e.g. pitch, duration, starting and ending times of each note, can be considered. According to [5][6][8], two features, rhythm and melody, are most useful in content-based music retrieval. Music with the same style often exhibits similar rhythm and melody [13]. Therefore, we adopt them as two representations of music data in this paper. For each of them, we derive the repeating patterns of each music piece. A repeating pattern [9] refers to a consecutive sequence of feature values that appear frequently in a music piece. It is generally agreed in musicology that the repeating pattern is one of the most important features in music representations. In this paper, we make repeating patterns useful for music classification by further incorporating constraints (i.e. length and frequency) to the repeating patterns. The repeating patterns that satisfy the constraints are called significant repeating patterns. Experiments on data sets of 1,000 song clips and 14 full-track real-world songs showed that this method can be successfully applied for music/voice separation, Competing with two recent state-of-the-art approaches. Further experiments showed that REPET can also be used as a preprocessor to pitch detection algorithms to improve melody extraction. After synthesizing the noise its being compressed.

### 1.1 Music/Voice Separation

Given a collection of MIDI files, we first select a representative track for each music piece manually. After that, the feature values of melody and rhythm are extracted from the representative tracks by using a MIDI parser. As a result, we represent each music piece by two symbolic sequences as follows. Rhythm stands for a sequence of beats in music and often brings people various kinds of perception. For example, a rhythm with fast tempos may make some people nervous but others excited. According to the duration of a note, we classify each note into one of the nine types in rhythm, where each type is notated as a distinct symbol called beat symbol. Following table shows the set of beat symbols we use in this paper. but for symbol I, the range of each beat symbol covers a quarter of a beat. The rhythm of a music piece can be Represented by a sequence of beat symbols, called the rhythmic sequence..

| Symbol | Duration  | Symbol | Duration  | Symbol | Duration      |
|--------|-----------|--------|-----------|--------|---------------|
| A      | (0,1/4)   | B      | (1/4,2/4) | C      | (2/4,3/4)     |
| D      | (3/4,4/4) | E      | (4/4,5/4) | F      | (5/4,6/4)     |
| G      | (6/4,7/4) | H      | (7/4,8/4) | I      | Above 2 beats |

**Table-1 Sets of beats symbols**

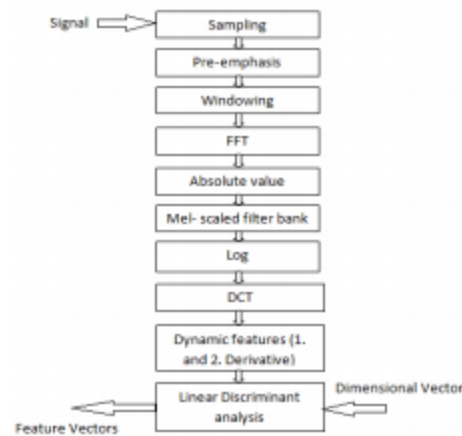
Melody is a sequence of pitches in music. A music piece with certain styles often contains specific melodies because the composer is used to showing a style by using similar melodies. A pitch interval stands for the difference between the pitch values of two consecutive notes. It is straightforward to transform a melody into a sequence of pitch intervals. According to the length of a pitch interval, we classify each pitch interval into one of the thirteen types in melody, where each type is notated as a distinct symbol called pitch symbol. Table 2 shows the set of pitch symbols we use in this paper. Each type of pitch intervals has two orientations,i.e.from low to high and the inverse. Therefore, we provide a plus or minus sign for each pitch symbol to indicate the orientation. In the set of pitch symbols, we distinguish the major intervals from the minor ones because they often bring people different kinds of perception, e.g.happiness and sadness. In this way, the melody of a music piece can be represented by a sequence of pitch symbols, called the melodic sequence.

| Symbol | Pitch interval | Symbol | Pitch interval | Symbol | Pitch interval | Symbol | Pitch interval |
|--------|----------------|--------|----------------|--------|----------------|--------|----------------|
| A      | 0              | B      | 2              | C      | 4              | D      | 5              |
| E      | 7              | F      | 9              | G      | 11             | H      | other          |
| A      |                | b      |                | d      |                | e      |                |
| F      | 10             | +      | Up             | -      | Down           |        |                |

**Table-2 set of pitch symbols**

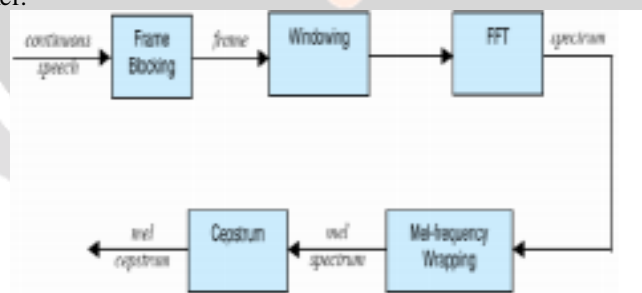
### 1.2 Mel Frequency Cepstrum Coefficient

we are using the Mel Frequency Cepstral Coefficients (MFCC) technique to extract features from the speech signal and compare the unknown speaker with the exits speaker in the database. Figure 7 shows the complete pipeline of Mel Frequency Cepstral Coefficients..



**Fig -1** Pipeline of MFCC

The Mel-frequency Cepstrum Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. Studies have shown that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency,  $f$ , measured in Hz, a subjective pitch is measured on a scale called the Mel scale. The Mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels. The following formula is used to compute the Mels for a particular frequency:  $mel(f) = 2595 * \log_{10}(1 + f/700)$ . A block diagram of the MFCC processes is shown in Figure. The speech waveform is cropped to remove silence or acoustical interference that may be present in the beginning or end of the sound file. The windowing block minimizes the discontinuities of the signal by tapering the beginning and end of each frame to zero. The FFT block converts each frame from the time domain to the frequency domain. In the Mel-frequency wrapping block, the signal is plotted against the Mel spectrum to mimic human hearing. In the final step, the Cepstrum, the Mel-spectrum scale is converted back to standard frequency scale. This spectrum provides a good representation of the spectral properties of the signal which is key for representing and recognizing characteristics of the speaker.

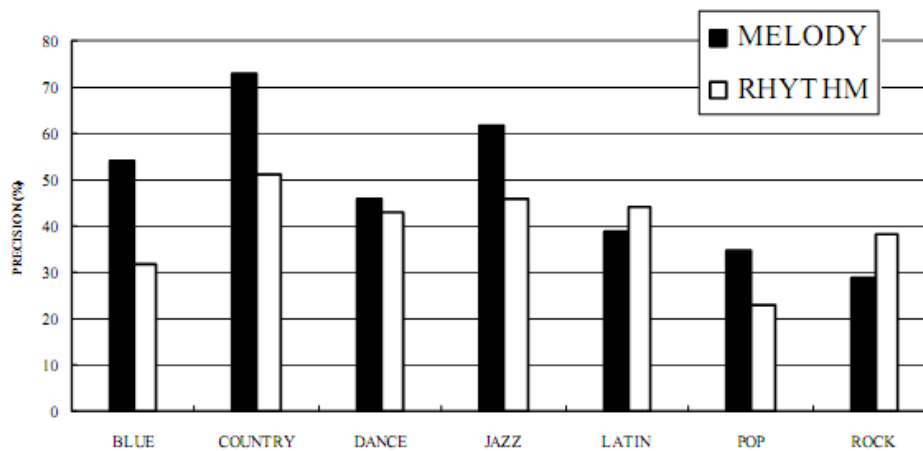


**Fig. 2** Block diagram of MFCC

## 2. Experiment Results

To evaluate the performance of our approach, we make a series of experiments to analyze the impacts of different features and thresholds. In addition, we also compare our approach with the one proposed by Chai and Vercoe [4]. In our experiments, we consider seven classes of music, including Blue, Country, Dance, Jazz, Latin, Pop, and Rock music. Furthermore, we select five hundred pieces of music from The New Zealand Digital Library [17] and then manually classify them based on the expertise collected from the World Wide Web. Each piece of music only belongs to one class. From these music, we select four fifth of them to derive the SRP's for training and utilize the others for testing. The precision and recall are computed as the averages of five different tests. The definitions of precision and recall are given as follows, where  $N_c$  is the number of correctly classified data,  $N_t$  is the number of testing data, and  $N_a$  is the minimum number of testing data that are required to make  $N$  data classified correctly.

$$\text{Precision} = N_c / N_t \quad \text{Recall} = N_c / N_a$$



**Fig. 3.** The precision for different features in the seven classes

In this experiment, we examine the influence of features on the precision of our approach With respect to the individual classes. According to the previous trials, we set the minimum constraint on frequency to 3 for Rhythm and 2 for melody, and the constraints on sequence length from 4 to 16. The experimental Results are shown in Figure 4, where three classes “COUNTRY”, “JAZZ”, and “BLUE” have the best precision (over 50%) for melody. The reason is because Music in these classes often contains particular melodies. On the Other hand, only two classes “ROCK” and “LATIN” Have better precision for rhythm than for melody. The reason is because music in these classes often impresses people a strong sense of rhythm. The class “POP” Has the worst precision for rhythm because it includes various kinds of music with different tempos.

### 3. Evaluation

Recently, FitzGerald et al proposed the Multipass Median Filtering based Separation (MMFS) method, a rather simple and novel approach for music/voice separation. Their approach is based on a median filtering of the spectrogram at different frequency resolutions, in such a way that the harmonic and percussive elements of the accompaniment can be smoothed out, leaving out the vocals. To evaluate their method, they fortunately found recordings released by the pop band The Beach Boys, where some of the complete original accompaniments and vocals were made available as split stereo tracks 3 and separated tracks 4. After resynchronizing the accompaniments and vocals for the latter case, we created a total of 14 sources in the form of split stereo wave files sampled at 44.1 kHz, with the complete accompaniment and vocals on the left and right channels, respectively. First, we compared the results of REPET with binary mask vs. soft mask, and without high-pass vs. with high-pass. A (non-parametric) Kruskal-Wallis one-way analysis of variance showed that using high-pass at 100 Hz on the voice estimates gave overall statistically better results, except for the voice SAR. Furthermore, using a soft mask gave overall slightly better results, except for the voice SIR. The improvement was however statistically not significant, except for the voice SAR. We nevertheless believe that the estimates sound perceptually better when using a soft mask instead of a binary mask, therefore we decided to show the results only for the soft mask. Since FitzGerald et al did not mention which tracks they used and only provided mean values, we could not conduct a statistical analysis to compare the results. We can however compare their means with our means and standard deviations, in the form of error bars. Thus, Fig. 2 and 3 show the average SDR, SIR, and SAR for the music and the voice estimates, respectively, at voice-to-music ratios of -6, 0, and 6 dB, without and with High-Pass at 100 Hz. The means and standard deviations of REPET are represented by the error bars and the means of MMFS are represented by the crosses.

### 4. CONCLUSIONS

In this paper, we propose a novel method for classifying music data by contents. We respectively extract rhythm and melody from music data and adapt the methods of finding repeating patterns to the needs of music classification. Given a music piece, we present a scheme for generating significant repeating patterns. A way to estimate the usefulness of SRP for classification is also proposed. For the music to be classified, we incorporate human perception and musicology into the similarity measures for SRP matching. Finally, we provide a complete procedure for determining which class a music piece should be assigned to. The experiment results indicate that some classes achieve better precision for a particular feature. Moreover, our approach performs on average better than the HMM-based approach. Experiments on a data set of 1,000 song clips showed that REPET can be efficiently applied for

music/voice separation, competing with two state-of-the-art approaches, while still showing room for improvement. More experiments on a data set of 14 full-track real-world songs showed that REPET is robust to real-world recordings and can be easily extended to full-track songs. Further experiments showed that REPET can also be used as a preprocessor to pitch detection algorithms to improve melody extraction. We have presented the REpeating Pattern Extraction Technique (REPET), a novel and simple approach for separating the repeating background from the non-repeating foreground in a mixture. The basic idea is to identify the periodically repeating segments in the audio, compare them to a repeating segment model derived from them, and extract the repeating patterns via time-frequency masking.

## 5. ACKNOWLEDGEMENT

The authors would like to thank C.-L. Hsu for providing the results of his singing voice separation system, J.-L. Durrieu for helping with the code for his music/voice separation system, and A. Klapuri for providing the code for his multiple estimator. We also would like to thank A. Liutkus and his colleagues from Telecom Paristech for their fruitful discussions, and our colleagues from the Interactive Audio Lab, M. Cartwright, Z. Duan, J. Han, and D. Little for their thoughtful comments. Finally, we would like to thank the reviewers for their helpful reviews.

## 6. REFERENCES

- [1] R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proceedings of IEEE Conference on Data Engineering, pp: 3-14, 1995.
- [2] C. Anagnostopoulou and G. Westermann, "Classification in Music: A Computational Model for Paradigmatic Analysis," Proceedings of the International Computer Music Conference, 1997.
- [3] J. J. Aucouturier and F. Pachet, "Music Similarity Measures: What's the Use?" Proceedings of International Symposium on Music Information Retrieval, 2002.
- [4] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models," Proceedings of International Conference on Artificial Intelligence, 2001.
- [5] C. C. Chen and Arbee L.P. Chen, "Query by Rhythm: An Approach for Song Retrieval in Music Database," Proceedings of IEEE Workshop on Research Issues in Data Engineering, pp: 139-146, 1998.
- [6] R. B. Dannenberg, B. Thom, and D. Watson, "A Machine Learning Approach to Musical Style Recognition," Proceedings of International Computer Music Conference, 1997.
- [7] S. Downie and M. Nelson, "Evaluation of a Simple and Effective Music Information Retrieval Method," Proceedings of ACM SIGIR Conference, pp: 73-80, 2000.
- [8] A. Ghias, H. Logan, D. Chamberlin, and B.C. Smith, "Query by Humming: Music Information Retrieval in an Audio Database," Proceedings of ACM Conference on Multimedia, pp: 231-236, 1995.
- [9] J. L. Hsu, C. C. Liu, and Arbee L.P. Chen, "Discovering Nontrivial Repeating Patterns in Music Data," Proceedings of IEEE Transactions on Multimedia, pp: 311-325, 2001.
- [10] S. Moshe, Dynamic Programming, Marcel Dekker Inc., 1992. [11] J. Paulus and A. Klapuri, "Measuring the Similarity of Rhythmic Patterns," Proceedings of International Symposium on Music Information Retrieval, 2002.
- [12] J. Pei, J. W. Han, B. Mortazavi-Asi, and H. Pinto, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," Proceedings of IEEE Conference on Data Engineering, 2001.
- [13] D. Pogue and S. Speck, Classical Music for Dummies, IDG books worldwide Inc., 1999. [14] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch Histograms in Audio and Symbolic Music Information Retrieval," Proceedings of International Symposium on Music Information Retrieval, 2002.
- [15] G. Tzanetakis, G. Essl, and P. Cook, "Automatic Musical Genre Classification Of Audio Signals," Proceedings of International Symposium on Music Information Retrieval, 2001.
- [16] B. Whitman and P. Smaragdis, "Combining Musical and Cultural Features for Intelligent Style Detection," Proceedings of International Symposium on Music Information Retrieval, 2002.
- [17] I. Witten (project leader) et al., "The New Zealand Digital Library Project," <http://nzdl2.cs.waikato.ac.nz/>, University of Waikato, New Zealand, April 2000.
- [18]. H. Schenker, Harmony. Chicago, IL: Univ. of Chicago Press, 1954.
- [19]. N. Ruwet and M. Everist, "Methods of analysis in musicology," Music Anal., vol. 6, no. 1/2, pp. 3-9+11-36, Mar.-Jul. 1987.
- [20]. A. Ockelford, Repetition in Music: Theoretical and Metatheoretical Perspectives. Farnham, U.K.: Ashgate, 2005, vol. 13, Royal Musical Association Monographs.
- [21]. J. Foote, "Visualizing music and audio using self-similarity," in Proc. 7th ACM Int. Conf. Multimedia (Part 1), Orlando, FL, Oct.-Nov. 30-05, 1999, pp. 77-80.