



## Signal Processing for Music Analysis

Poonam Priyadarshini

Department of Electronics and Communication

Birla Institute of Technology Mesra, B.I.T. Patna, Bihar, INDIA

Soubhik Chakraborty

Department of Applied Mathematics, Birla Institute of Technology,  
Mesra, Ranchi, Jharkhand, INDIA

**Abstract:** Music information retrieval (MIR) is an emerging research area for both the research community and music industry. It retrieves music by querying or humming from large music database. People searched music of their choice from database by song title, composer and performer. Features of interest may include melody, harmony, rhythm, and instrumentation. Classification of music is done as mood classification, artist recognition, instrument recognition, etc. This paper discuss the different methods of features extraction of music signal. In this paper we are using Indian music Raga as a database.

**Keywords:** Music information retrieval (MIR), Linear Predictive Coding coefficients (LPC), Mel frequency cepstral coefficient (MFCC).

### I. INTRODUCTION

WITH the development of information and multimedia technologies, digital music are available from different media, like radio broadcasting, digital storage such as compact discs (CDs), the Internet, etc. Music information retrieval (MIR) is an emerging research area in multimedia to cope with such necessity. A key problem in MIR is classification, which differentiate each song based on genre, mood, artists, etc. Music classification is an interesting topic because most end users may only be interested in certain types of music. Music listeners often wish to access and use their music collections according to salient features of music in the audio recording. Features of interest of listeners may include melody, harmony, rhythm, and instrumentation in music. The user can search for the music they are interested from the classification system. On the other hand, different category of music have different properties. We can manage them more effectively and efficiently once they are categorized into different groups.

Most existing work in music information retrieval analyzes music via low-level features such as Mel frequency cepstral coefficient (MFCC) and other spectral coefficients. However, low-level features are insufficient for many applications since they are related to the signal characteristics rather than the semantic content of music. On the contrary, mid-level features such as chord, rhythm, and instrumentation represent musical attributes and contain rich information for music analysis.

One of the most important mid-level features of music is chord sequence, which describes harmonic progression and tonal structure of music. Since harmonic progression is strongly related to the perceived emotion, similar chord sequences can be observed in songs that are close in genre, emotion, etc. With chord sequence, songs that are similar in various aspects can be identified and retrieved more effectively.

At the same time, content-based retrieval technology is developing as hot topics, such as content-based image retrieval, content-based audio/video retrieval. Compared with traditional keyword-based music retrieval, content-based music retrieval provides more flexibility and expressiveness. Users may express queries by humming, singing and whistling song.

Low-level features are the important building block for audio classification systems. They are easy to extract and have demonstrated good performance in virtually all music classification tasks. Low-level features can be further divided into two classes of timbre and temporal features.

Timbre describe the quality of a sound, different timbres are produced by different types of sound sources, like different voices and musical instruments. The extraction of timbre features is closely related to spectral analysis of the audio signal. A song is usually split into many local frames in the first step to facilitate subsequent frame level timbre feature extraction. It has two advantages. Firstly, it leads to a more efficient feature extraction process, since we only need to apply spectral analysis to short-time signals. Secondly, it is more effective to model the timbre features in frames with 10–100 ms duration. Due to the non-stationary nature of music, the statistical properties of audio signal of a song depend on time.

By applying frame-level feature analysis, we can assume that the audio signal within each frame is stationary. After framing, spectral analysis techniques such as fast Fourier transform (FFT) are then applied to the

windowed signal in each local frame. From the output magnitude spectra, we find features such as spectral centroid (SC), spectral roll off (SR) and spectral bandwidth (SB) capturing statistics of the spectra. For MFCC, the sub bands are linearly spaced in lower frequencies (less than 1KHZ) and logarithmically spaced in higher frequencies. The reason for using log-scale at the higher frequency range is to mirror the human auditory system, whereby finer spectral resolution is achieved at lower frequencies. The local windows used in timbre feature extraction are usually taken from fixed-size intervals regardless of the underlying musical events. It would be desirable to split the signal into unequal intervals corresponding to separate notes where the boundaries are aligned with the onset and offset of the notes [8]. Secondly, the bank of band-pass filters is usually organized to cover octave-based logarithmic spectral range.

**Temporal features** form another important class of low-level features that capture the temporal evolution of the signal. It is worth noticing that temporal features are usually constructed on top of timbre features or the spectrogram obtained during timbre feature extraction. Timbre features extracted from several frames are further integrated to create the temporal feature. In this way, we can create a richer set of features for classification with different combination of timbre and temporal features. The simplest type of temporal features is statistical moments such as mean, variance, covariance, and kurtosis of timbre features collected from a larger local window called texture window [9].

#### **Mid-Level Features**

Despite the good performance of low-level features for music classification, they do not capture the intrinsic properties of music that humans perceive and appreciate. The commonly used mid-level features in music analysis include

- Rhythm—recurring pattern of tension and release in music;
- Pitch—perceived fundamental frequency of the sound;
- Harmony—combination of notes simultaneously, to produce chords, and successively, to produce chord progressions.

Rhythm describes how certain patterns occur and recur in the music. Beat and tempo (beat-per-minute, BPM) are two important cues that describe rhythmic content of the music which have been utilized in music classification.

#### **Features Extraction**

Input query to any information retrieval system consists of relevant and irrelevant information. Features Extraction is the process of discarding as much irrelevant information as possible and representing relevant information in a compact and meaningful form. During the process of proposing an MIR system we investigated a variety of features including some statistical features such as statistical mean, standard deviation and some basic signals properties such as zero-crossing rate and energy. Moreover, we applied some of the digital signal processing algorithms to musical data to extract features such as Fast Fourier Transform FFT, Linear Predictive Coding coefficients LPCC and MFCC. The algorithm for getting MFCC feature is given below.

1. pre-emphasis 2. Framing 3. Hamming windowing 3. Fast fourier Transform to obtain power spectrum 4. Log of FFT 5. Mel filter bank 6. DCT for decorrelation 7. delta delta MFCC coefficients.

## **II. TECHNIQUES OF FEATURE EXTRACTION**

The methodology of audio classification involves extracting discriminatory features from the audio data and feeding them to a pattern classifier. Few techniques generate a pattern from the features and use it for Classification by the degree of correlation. Few other techniques use the numerical values of the features coupled to statistical classification method. In general, methods for detecting pitches can be divided roughly into two categories: time-domain based and frequency-domain based.

In the time-domain, Zero Crossing Rate (ZCR) is a popular method. The basic idea is that ZCR gives information about the spectral content waveform cross zero per unit time. Recently, ZCR appeared in a different form such as VZCR (variance of ZCR) or SZCR (smoothing ZCR) (Huang and Hansen, 2006). They are more suitable for dealing with audio files than ZCR.

In the frequency-domain, Fast Fourier Transformation (FFT) is one of the most popular methods. This method is based on the property that every waveform can be divided into simple sine waves. But, a low spectrum rate for longer window may increase the frequency resolution while decreasing the time resolution. Another problem is that the frequency bins of the standard FFT are linearly spaced, while musical pitches are better mapped on a logarithmic scale. The other methods are

MFCC (Mel frequency cepstral co-efficient) LPC (Linear predictive coding) LDB (local discriminant bases).

**Classifiers for Music Classification:** In standard classification, we are presented with a training data set where each example comes with a label. The purpose is to design a classification rule that can best predict the labels for unseen data.

Classifier design is a standard topic in pattern classification. The common choices of classifiers are K-nearest neighbour (K-NN) support vector machine (SVM) and GMM classifier. Various other classifiers have also been used for different music classification tasks, including logistic regression artificial neural networks (ANN) [7].

8] decision trees[8] linear discriminant analysis (LDA)[6] ,nearest centroid (NC and sparse representation-based classifier (SRC) K-NN and SVM are the two most popular classifiers used for both general classification problems and in music classification as well. K-NN uses training data directly for the classification of testing data. The label of the testing instance is predicted by majority voting on the labels of the nearest instances in the training set. SVM is the state-of-the-art binary classifier based on the large margin principle. Given labeled instances from two classes,

Therefore, the SVM has good classification performance since it focuses on the difficult instances. Both K-NN and SVM are applicable to single feature vector representations and pairwise similarity values as well. In the latter case, a kernel matrix is built from pairwise similarity values that can be used directly by the SVM .

Another classifier that can directly handle feature set classification is convolutional neural network (CNN) which is a generalization of the standard neural network model by taking convolutions over the segments of the input signal. Hence, such model can be used for audio classification based on sequence of timbre features like raw MFCC features. MFCC is inarguably the single most important feature type that was widely used for genre classification. It alone can produce quite good classification result compared to many other methods that use more sophisticated collections of features.

#### B. Statistical Features

1. **Skewness** : Skewness is the measure of the symmetry or asymmetry of a signal.

$$\text{Skewness} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3}{(N-1)s^3}$$

Y is the mean, s is the standard deviation, and N is the number of data points.

2. **Kurtosis**: Kurtosis is a measure of the noisiness of a signal. It shows whether the data is peaked or flat relative to a normal distribution.

$$\text{Kurtosis} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)s^4}$$

**Experiment:** We are choosing Indian raga for e.g. Todi Raga Bihag Raga for doing analysis of this music signal.

**Result:** The results of the experiment are shown below.

#### REFERENCES

- [1] R. Typke, F. Wiering and R.C. Veltkamp "A survey of Music Information Retrieval Systems," in Proc. Int. Conf. Music Inf. Retrieval (ISMIR), 2005, pp. 153-160.
- [2] ISMIR: The international conferences on music information retrieval and related activities [Online]. Available: <http://www.ismir.net/>
- [3] J. S. Downie, "Music information retrieval," in *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, 2003, vol. 37, ch. 7, pp. 295-340.
- [4] P. Herrera, X. Amatriain, E. Battle, and X. Serra. Toward instrument segmentation for music content description: a critical review of instrument classification techniques. presented at Int. Symp. Music Information Retrieval (ISMIR 2000). [Online]. Available: <http://ismir2000.indiana.edu/>
- [5] A.P.Klapuri, A. J.Eronen, J.T.Astola, "Analysis of the Meter of Acoustic Musical Signals," IEEE Trans. Audio, Speech, and Language Processing, 2006, 342-355.
- [6] T. Li, M. Ogihara, and Q. Li, "A comparative study of content-based music genre classification," in *Proc. SIGIR*, 2003
- [7] A. Meng and J. Shawe-Taylor, "An investigation of feature models for music genre classification using the support vector classifier," in *Proc. Int. Conf. Music Information Retrieval*, 2005.
- [8] I. Mierswa and K. Morik, "Automatic feature extraction for classifying audio data," *Mach. Learn.*, vol. 58, pp. 127-149, 2005..
- [9] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293-302,
- [10] K.West and S. Cox, "Finding an optimal segmentation for audio genre classification," in *Proc. Int. Conf. Music Information Retrieval*, 2005.2002
- [11] Y. E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features," in *Proc. Int. Conf. Music Information Retrieval*, 2002.
- [12] F. Morchen, A. Ultsch, M. Thies, and I. Lohken, "Modeling timbre distance with temporal statistics from polyphonic music," *IEEE Trans.Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 81-90, 2006.
- [13] M. Mandel and D. Ellis, "Song-level features and SVMs for music classification," in *Proc. Int. Conf. Music Information Retrieval*, 2005.

#### Skewness

TODI RAAG 1st segment .The Spectral skewness related = 4.9419

TODI RAAG 2nd segment The Spectral skewness related = 5.9917

TODI RAAG 3rd segment.The Spectral skewness related = 5.0561

#### Kutosis

TODI RAAG 1st segment.The Spectral kurtosis = 31.833

TODI RAAG 2nd segment.The Spectral kurtosis = 56.2213

TODI RAAG 3rd segment.The Spectral kurtosis = 32.0919

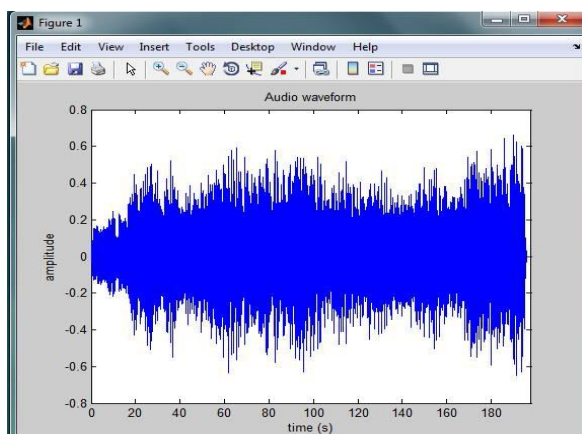


**Figure 1: Reading of audio sample**

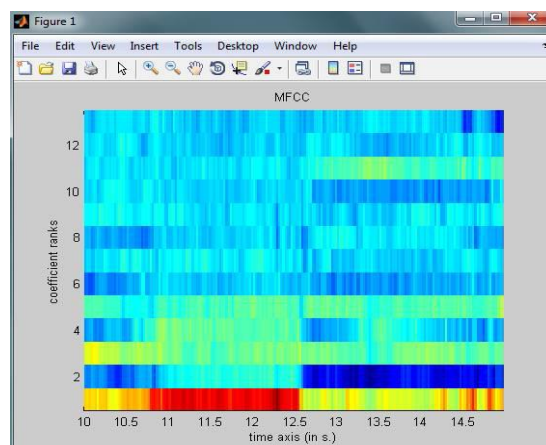


**Figure 3: Arrangement of frequencies into Mel bands**

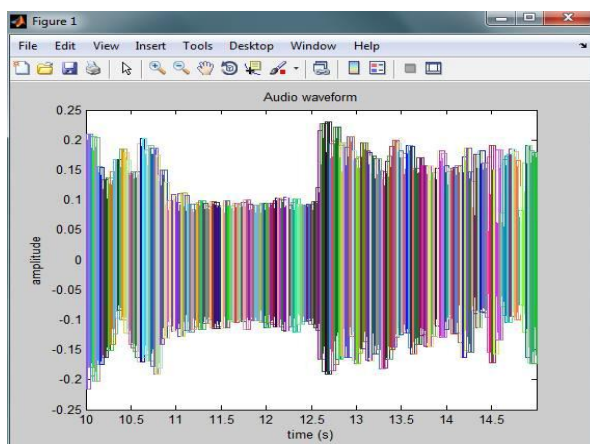


**Figure 2: Framing of segment**

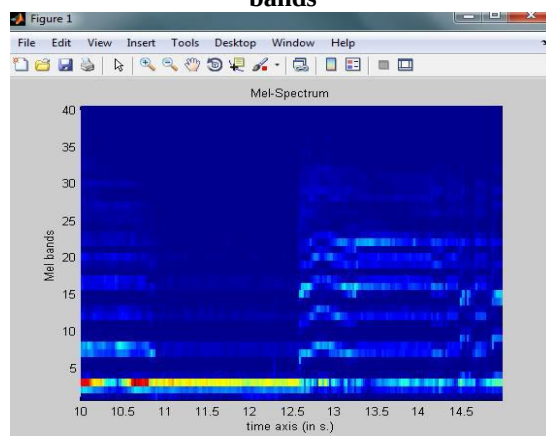


**Figure 4: Conversion from frequency domain to Mel spectrum domain**

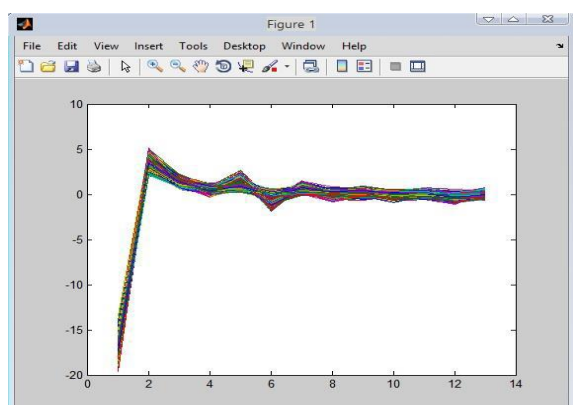


**Figure 5: mfcc plots of todi segment 1**

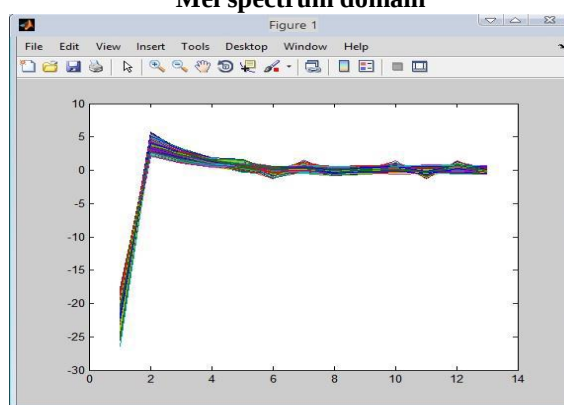


**Figure 6: mfcc plots of todi segment 2**

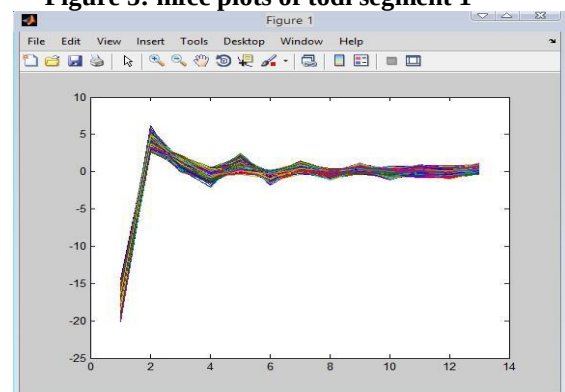


**Figure 7: mfcc plots of Bihag segment 1**

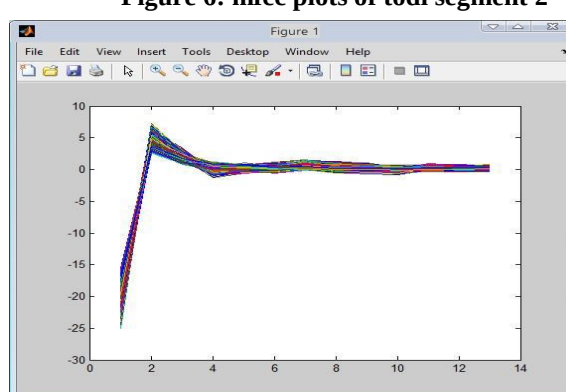


**Figure 8: mfcc plots of Bihag segment 2**