PAN
ARCHIVES OF ACOUSTICS
Vol. **42**, No. 2, pp. 213–222 (2017)

DOI: 10.1515/aoa-2017-0024

Automatic Genre Classification Using Fractional Fourier Transform Based Mel Frequency Cepstral Coefficient and Timbral Features

Daulappa Guranna BHALKE, Betsy RAJESH, Dattatraya Shankar BORMANE

Deptartment of Electronics & Telecommunication
JSPM's Rajarshi Shahu College of Engineering
SPPU, Pune, India; e-mail: bhalkedg2000@yahoo.co.in, betsyrajesh@gmail.com, bdattatraya@yahoo.com

(*received October 12, 2015; accepted February 1, 2017*)

This paper presents the Automatic Genre Classification of Indian Tamil Music and Western Music using Timbral and Fractional Fourier Transform (FrFT) based Mel Frequency Cepstral Coefficient (MFCC) features. The classifier model for the proposed system has been built using K-NN (K-Nearest Neighbours) and Support Vector Machine (SVM). In this work, the performance of various features extracted from music excerpts has been analysed, to identify the appropriate feature descriptors for the two major genres of Indian Tamil music, namely Classical music (Carnatic based devotional hymn compositions) & Folk music and for western genres of Rock and Classical music from the GTZAN dataset. The results for Tamil music have shown that the feature combination of Spectral Roll off, Spectral Flux, Spectral Skewness and Spectral Kurtosis, combined with Fractional MFCC features, outperforms all other feature combinations, to yield a higher classification accuracy of 96.05%, as compared to the accuracy of 84.21% with conventional MFCC. It has also been observed that the FrFT based MFCC effieciently classifies the two western genres of Rock and Classical music from the GTZAN dataset with a higher classification accuracy of 96.25% as compared to the classification accuracy of 80% with MFCC.

Keywords: feature extraction; Timbral features; MFCC; Fractional Fourier Transform (FrFT); Fractional MFCC; Tamil Carnatic music.

1. Introduction

Digital technology has completely restructured the music industry, because of which consumers have access to thousands of music tracks stored locally on their smartphones and millions of records instantly available through cloud-based music services. Recent technological advances help users interact with music by directly analyzing the musical content of audio files (SHAO *el al.*, 2005). The vast amount of available musical databases creates the need for reliable methods of searching and organizing them, and demands novel methods of description, indexing, searching, and interaction (BENETOS, KOTROPOULOS, 2010; SCARINGELLA *et al.*, 2006). MIR (Music Information Retrieval) deals with the automatic analysis of music signals and uses various characteristics that best describe the music content (SCARINGELLA *et al.*, 2006). Genre information is one such characteristic that can help describe music content and is a fundamental component of MIR (FU *et al.*, 2011).

The music of India has very ancient roots and has existed for many millennia. Different musical forms like the North Indian Hindustani, South Indian Carnatic, Ghazals, diverse forms of folk music, film music and Indo-western fusion music contribute to the Indian music (BAGUL *et al.*, 2014). The Tamil language of Tamil Nadu, South India, has an antiquity going beyond the period of at least B.C. 1250 (Tamil Music, 2011). Tamil music is classified into two main genres, namely (1) **Tamil Classical** – structured music, sung to a rhythmic cycle or tala (called the *Carnatic style music*) (KUMAR *et al.*, 2014) and (2) **Tamil Folk** – *rural* music composed in colloquial style. Most devotional songs called as 'Keeerthanai' or 'Kriti' and are composed in classical style. The 'Keertanai' consist of (1) 'Pallavi' (first section of the song indicating the theme), (2) 'Anupallavi' (forms the chorus along with Pallavi, to be repeated) and (3) 'Charanams' (foot of the song) (ASHOK NARAYANAN, PRABHU, 2003). In this work, various original compositions of Tamil 'gospel Keertanai' and patriotic compositions have been employed for the purpose of classification. These devotional and patriotic musical compositions have been composed on classical ragas (melodies) and tala (rhythm), mostly sung as a solo (Vedanayagam Sastriar, 2011).

Tamil folk music is known for the tala intricacies and the ancient folk music was based on classical ragas or melodies like Manji, Sama, Navaroz, and Kalyani. Instrument accompaniments are Nadhaswarams (a type of flute), drums or melam and cymbals or Kaimani. Folk music has continued to evolve over the years. The present day folk music can be classified as "Naatupurapaadalgal' (rural folk music) and "Gaana padalgal' (urban folk music). The following section discusses the various methodologies that have been employed in the previous years for automatic genre classification of both western and Indian music.

1.1. Literature survey

Tzanetakis, Cook (2002) have done pioneering work in automated genre classification of western music and deduced three essential features for musical content namely timbral texture, rhythm and pitch content features. Statistical pattern recognition classifiers have been trained for real-world music data. Frames and whole songs were used for which, a classification accuracy of 61% has been achieved for ten western musical genres. The results closely matched the reported results for human musical genre classification. Further, Li and Tzanetakis (2003) have described the factors in Automatic Genre Classification and studied the performance of Support Vector Machines and LDA (Linear Discriminant Analysis) classifiers. LDA was used to find discriminative feature transform as eigenvectors, to capture both intraclass and interclass separation.

Meng *et al.* (2007) proposed the temporal integration of short time features using Multivariate autoregressive models (MAR). The idea has been to extract a summarized power of each feature dimension independently in four specified frequency bands. MAR has been used for temporal feature integration since it has the potential of modeling both temporal correlations and dependencies among features. Li *et al.* (2010) extracted features from Daubechies wavelet coefficients histogram (DWCH) and have observed that timbral features combined with MFCC yield high accuracy. To extract more powerful features like DWCH and OSC (Octave-based Spectral Contrast) subband analysis has been performed where the power spectrum was decomposed into subbands and features were extracted from each subband.

Lim *et al.* (2012) proposed a Music-Genre Classification System based on Spectro-Temporal Features and Feature Selection. The mean, variance, minimum and maximum values, spectral modulation flatness, crest, contrast and valley features were estimated and Support Vector Machine (SVM) was used as a classifier. The method has proved to have higher accuracy at a lower feature dimension for the GTZAN and ISMIR2004 databases.

Chen *et al.* (2012) improved classification accuracy using wavelet package transform (WPT), since WPT performs a wavelet decomposition that offers a richer signal analysis. A best basis algorithm selection has been performed using top- down search strategy. Mel-frequency Cepstral Coefficients (MFCC) and log energies extracted from the decomposition coefficients were used to build the SVM classifier with resulting accuracy of 89.03%. Baniya *et al.* (2014) derived a feature set that included higher order moments of skewness, kurtosis and covariance of features in addition to their mean and variances, resulting in improvement of classification accuracy to 85.15%. Rosner *et al.* (2014) proposed the classification of genresbased on Music Separation into Harmonic and Drum Components. They have employed co-training (semi-supervised learning) to SVM-based classification, which enabled the SVM to learn from a small training set, which later helped to classify unlabelled data iteratively.

Nagavi *et al.* (2011), in the overview of classification and retrieval systems of Indian music, have verified that PCD (Pitch class Distribution), tone profiles and spectral profiles sufficiently discriminate ragas automatically. The pitch class or chroma is a representation of pitches from all octaves, mapped to a single octave.

Kini *et al.* (2011) have classified bhajan and qawwali sub-genres of North Indian devotional music with timbral features, tempo and modulation spectra of timbral features. They achieved 92% accuracy by applying 10 fold cross validation of tempo estimations, feature summaries of mean-variance and envelope modulation with Support Vector Machine (SVM) and Gaussian Mixture Model (GMM). Rao (2012) worked on the extraction of metadata for Hindustani Classical music using factual information that accompanies music on a CD, such ascomposer, genre, artist and other semantic labels such as mood. Pitch detection, rhythm detection and melody estimation through motif (repetitive phrase) identification and oscillation (gamakas), were employed for classification of ragas of Hindustani music.

Bhalke *et al.* (2015) had proposed Fractional Fourier Transform (FrFT) based MFCC features for discriminating musical instruments and have found that the interclass variation was greatly maximised and intra-class variation was minimised by the use of the chirp like kernel basis function of the FrFT.

Various features such as Timbral, Temporal, Spectral, Wavelet and MFCC have been proposed in the past years for music genre classification. Also, it has been observed that Timbral and MFCC features have made a significant contribution to genre classification (Bhalke *et al.*, 2014; Fu *et al.*, 2011; Ghosal *et al.*, 2012). In addition, Fractional Fourier Transform makes use of chirp decomposition which is especially suitable for music signals in which time, frequency and phase information play a vital role (Ashok Narayanan, Prabhu, 2003; Bhalke *et al.*, 2016).

It has further been observed, that very little work has been done on Indian genres and there has been no previous work on South Indian Tamil music. Thus, the next sections of the paper present a novel feature extraction scheme for automatic music genre classification of Indian Tamil music, that include FrFT based MFCC features or Fractional MFCC (FrMFCC) using two classifiers namely KNN and SVM. The comparative results of the two classifiers with the proposed and previous methods have been tabulated and presented. Also the results have been tested for western music genres too.

The sections that follow include proposed system, feature extraction and database details in Sec. 2, experimental results in Sec. 3, conclusion and future work in Sec. 4, acknowledgements in Sec. 5 and finally the list of references in Sec. 6.

2. Proposed system and feature extraction

Various methodologies have been used for Automatic genre classification. The two major stages in genre classification are (1) Feature extraction and (2) Classification. The first stage is to extract the meaningful and relevant features from audio that could sufficiently discriminate the music genres. The second step is training a suitable classifier with extracted feature values and then testing it with new samples.

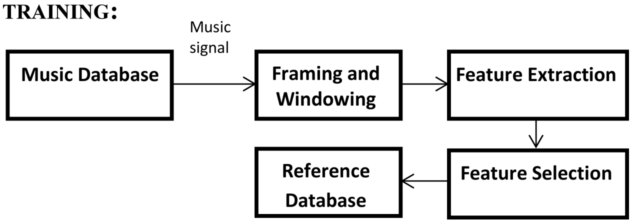

Fig. 1. Training phase.

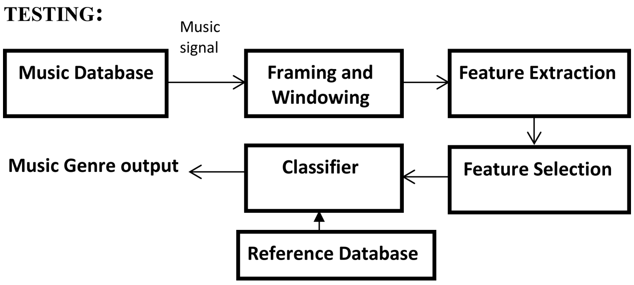

Fig. 2. Testing phase.

2.1. Framing and segmentation

To be able to discriminate sufficiently between the genres, 30-second clips were taken from the songs. It is assumed that any randomly occurring signal is stationary, and thus the properties remain invariant for 10 ms to 20 ms. This assumption makes it possible for signal processing techniques to apply to the short stationary signals. Thus, the 30-second excerpts were further framed into 20 ms frames with 50% overlapping.

2.2. Windowing

To maintain continuity of the first and last points in the frame, a Hamming window is multiplied with each frame. Since, the Hamming window is a smooth window and reduces the size of the side lobes, it has been employed for windowing. For a signal frame given by $X_s(n)$, where $n = 0, 1, \ldots, N-1$, the windowed signal is given by $X_s(n) * W(n)$. The Hamming window $W(n)$ is described as

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right). \tag{1}$$

2.3. Feature extraction

Meaningful and relevant features, that could sufficiently discriminate the music genre are extracted from the audio clippings in this phase. The timbral, rhythmic and pitch features extracted are as follows:

Table 1. List of features.

Feature number	Feature class	Number of features	Features used
1–6	Timbral (Spectral)	6	Mean & Std. Dev. of Spectral Centroid, Spectral Roll off, Spectral Flux
7–32	MFCC	26	Mean & Std. Dev. of MFCC features
33–40	Statistical	8	Mean & Std. Dev. of Spectral Skewness, Spectral Kurtosis, Flatness, Entropy
41–66	Fractional MFCC	26	Mean & Std Dev. of Fractional MFCC
67–68	Temporal	2	Mean of Zero Crossing Rate and Root Mean Square Energy

2.3.1. Timbral features

The timbral features have been obtained from the frequency domain of the signal. The signal has been first transformed into the frequency domain and various spectral features have been extracted from the spectrum.

Spectral Centroid: The Spectral Centroid gives the measure of the brightness of a sound. The centroid of a spectral frame can be defined as the average frequency weighted by amplitudes, and dividing it by the sum of the amplitudes and can be given by

$$\text{Spectral Centroid} = \frac{\sum\limits_{k=1}^{N} kM[k]}{\sum\limits_{k=1}^{N} M[k]}, \tag{2}$$

where $M[k]$ is the magnitude of the FFT at frequency bin k and N is the number of frequency bins. Centroid finds this frequency for a given frame and then finds the nearest spectral bin for that frequency.

Spectral Roll-off: The Spectral Roll-off is a measure of the spectral shape. It is defined as the frequency bin M below which the 85% of the magnitude distribution is concentrated

$$\text{Roll off} = \sum_{k=1}^{M} M[k] = 0.85 \sum_{k=1}^{N} kM[k]. \tag{3}$$

Spectral Flux: The Spectral Flux gives the rate of change of the power spectrum and indicates how quickly the power changes from frame to frame. The spectral flux has been calculated by comparing the power spectrum of one frame with the power spectrum of the previous frame and is given by:

$$\mathbf{F} = ||M[k] - Mp[k]||, \tag{4}$$

where $Mp[k]$ denotes the FFT magnitude of the previous frame in time.

2.3.2. Mel-frequency Cepstral coefficients (MFCC)

Mel Frequency Cepstral Coefficient (MFCC) is short time power spectral representation of a signal. It provides useful information regarding psychoacoustic property of human auditory system. Block schematic of MFCC is shown in Fig. 3. This feature extraction consists of pre-processing, pre-emphasis, framing, windowing, FFT, Triangular mel band pass filter and DCT. In preprocessing the silence part of the signal is removed using ZCR and energy features. This helps to reduce the computational complexity of the system. pre-emphasis is done to boost the high frequency components using first order high pass filter. Framing is done with 20 ms duration with 10 ms overlapping. Then the signal is windowed with hamming window. Further, the signal is transformed into spectral domain and passed through 24 mel frequency triangular band pass filters. Log values of these spectral components have been obtained. Discrete Cosine Transform (DCT) of these log values have been taken to decorrelate the signal. Thirteen most significant MFCC coefficient have been obtained for each frame. Statistical values such as mean value of these coefficients have been computed and used as feature vector.

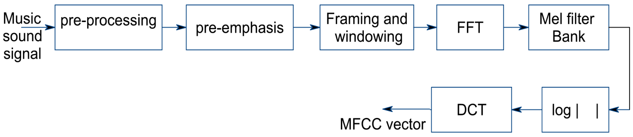

Fig. 3. Block schematic of Mel frequency cepstral coefficients.

2.3.3. FrFT based MFCC features

The time-frequency representation of a signal is on a plane with two orthogonal axes, where the time axis is represented horizontally as $x(t)$ and the frequency axis is represented vertically. The conventional Fourier Transform $X(\omega)$ of a signal $x(t)$ is represented along the frequency axis. The Fourier Transform $F[x(t)] = X(\omega)$, employs the Fourier Transform operator FT, which rotates the time axis anticlockwise by $\pi/2$ radians. The Fractional Fourier transform operator FrFT likewise rotates the signal by an angle that is a non-multiple of $\pi/2$ radians (Ashok Narayanan, Prabhu, 2003). Thus, the signal is represented in a plane that makes an angle 'α' to the time axis where $\alpha = a * \pi/2$. The value of 'a' can lie anywhere between 0 and 1. For this work, the value $a = 0.98$ was found to yield better results.

FrFT uses linear chirps as a basis function and thus, it offers a great deal of flexibility in the processing of audio signals (Bhalke *et al.*, 2016).

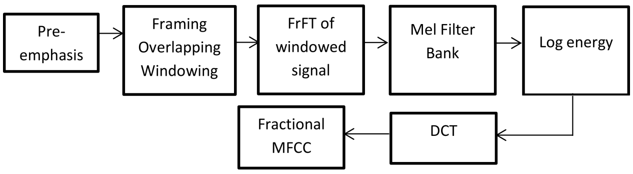

Fig. 4. Block schematics of Fractional Mel Frequency Cepstral Coefficients (FrMFCC).

FrFT is represented by F^{α}. FrFT has following properties:

(1) $F^0 = I$, zero rotation or identical operator corresponds to time domain,

(2) $F^{\pi/2} = F$, Corresponds to Fourier transform operator,

(3) $F^{2\pi} = I$, 2π rotation,

(4) $F^{\alpha}F^{\beta} = F^{\alpha+\beta}$, additivity of rotation.

FrFT of a signal $x(t)$ with order α is given by $F^{\alpha}(u)$ which is represented by

$$F^{\alpha}(u) = \int_{-\infty}^{\infty} x(t) K_{\propto}(t,u)\,\mathrm{d}t, \tag{5}$$

where $K_\alpha(t,u)$ is a transformation kernel and is given by Eq. (6)

$$K_\alpha(t,u) = \begin{cases} \left(\sqrt{\frac{1-j\cot\alpha}{2\pi}}\right) e^{j((t^2+u^2)/2)\cot\alpha - ju\,t\,\csc\alpha}, & \\ & \text{if } \alpha \text{ is not multiple of } 2\pi, \\ \delta(t-u), & \text{if } \alpha \text{ is multiple of } 2\pi, \\ \delta(t+u), & \text{if } \alpha+\pi \text{ is multiple of } 2\pi. \end{cases} \quad (6)$$

The FrFT based MFCC features are calculated by preprocessing, framing (with 50% overlap), windowing (Hamming window of size 1024), Fractional Fourier Transform, positioning on a Mel filter bank, calculation of Log energy, energy compaction by DCT and finally calculation of Fractional MFCC from the signal. The Mean and Standard Deviation of the FrMFCC values, along the frames, for a texture window is calculated and taken as feature vectors.

2.3.4. Temporal features

Temporal features give the evolution of the signal over a period of time. The temporal features that have been extracted are:

Zero Crossing Rate (ZCR): It is defined as the number of times a signal crosses the X-axis. ZCR gives an idea of the frequency of the signal and is given by:

$$\mathrm{ZCR} = \frac{1}{N}\sum_{0}^{N-1} \left|\mathrm{sgn}[m(n)] - \mathrm{sgn}[m(n-1)]\right|, \quad (7)$$

where N is the total number of samples, $m(n)$ and $m(n-1)$ is the signal at n-th and $(n-1)$-th sample respectively.

2.3.5. Energy features

Root Mean Square Energy: The global energy of the signal $x[n]$ has been computed simply by taking the root average of the square of the amplitude, also called root-mean-square energy (RMS)

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{n} x_i^2}, \quad (8)$$

where x_i is the amplitude of the signal.

2.3.6. Statistical features

Entropy: The Entropy gives a description of the input curve p and indicates whether it contains predominant peaks or not. It is calculated using Shannons entropy based on the equation:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_b p(x_i), \quad (9)$$

where p indicates the curve and b is the base of the algorithm.

Spectral skewness: The spectral skewness is the third central moment and is a measure of the symmetry of the distribution. A positive value indicates a positively skewed distribution with few values larger than the mean and thus has a long tail to the right. A negatively skewed distribution has a longer tail to the left. Skewness is given by

$$\mu = \int (x-\mu_1)^3 f(x)\,\mathrm{d}x, \quad (10)$$

where μ is the mean of the distribution.

Spectral kurtosis: The spectral kurtosis is defined as the fourth standardised moment and is defined as

$$\mathrm{kurtosis} = \frac{\mu_4}{\sigma^4}, \quad (11)$$

where μ is the mean and σ is the variance. Kurtosis gives the sharpness of the peaks.

Spectral flatness: The spectral flatness indicates whether the distribution is smooth or spiky. It is calculated as the ratio between the geometric mean and the arithmetic mean:

$$\mathrm{flatness} = \frac{\sqrt{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}}. \quad (12)$$

2.4. Classification

After the feature extraction, the summarised feature values were fed to classifiers for modeling and prediction. In this work, two classifiers were used: KNN (K-Nearest Neighbour) and SVM.

KNN (*K*-Nearest Neighbours): The KNN is a simple algorithm that stores all available class labels and decides the class of a test sample based on a similarity measure (e.g., distance functions) (Fu *et al.*, 2011). KNN is used in statistical estimation and pattern recognition as a nonparametric technique which does not make any assumptions about the underlying data distribution (Scaringella *et al.*, 2006). The KNN is a lazy learning algorithm that does not use the training samples to perform any generalisation and the entire data is used for the testing phase without discarding any. So in the KNN algorithm there is a minimal cost involved in the training but a high cost involved in testing both in terms of time and memory since all data points are utilized and stored for decision making. Various distance functions are used for measuring and are as follows;

$$\text{Euclidean:} \quad \sqrt{\sum_{i=1}^{k}(x_i-y_i)^2}, \quad (13)$$

$$\text{city block:} \quad \sum_{i=1}^{k}|x_i-y_i|, \quad (14)$$

$$\text{Chebychev:} \quad \max\{|x_i-y_i|\}, \quad (15)$$

where x_i and y_i are two instances and the distance between them is defined by $d(x_i, y_i)$. The standard Euclidean, city block and Chebychev distances are given by Eqs. (13), (14) and (15), respectively. The KNN classifier is a very simple algorithm that works well for real world data where the classes may be linearly separable or not. The value of K and the distance metric alone need to be tuned.

Support Vector Machines (SVM): The Support Vector Machine is a supervised learning algorithm that is used for classification and regression analysis. The SVM builds a model with a training set that is presented to it and assigns test samples based on the model. An SVM model represents points of samples in space, mapped in such a way that the samples of the separate categories are as wide as possible (Scaringella *et al.*, 2006) The challenge is to find the optimal hyperplane that maximises the gap. New samples are then mapped into that same space and class predictions are made as belonging to either category based on which side of the gap they fall on. The performance of the SVM is greatly dependent on its kernel functions (linear, polynomial or exponential). For the purpose of this experiment, the more popular Radial Basis Function (RBF) kernel has been chosen. RBF is a squared exponential kernel, capable of handling complex data and is more flexible since it gives access to all infinitely differentiable functions. The RBF kernel for two samples x and x' is defined by

$$K(x, x') = \exp\left(\frac{\|x - x'\|^2}{2\sigma^2}\right), \tag{16}$$

where $\|x - x'\|^2$ is the squared Euclidean distance between the feature vectors and σ is a free parameter and the parameter $\gamma = 1/2\sigma^2$.

2.5. Dataset

In this work, two types of databases were used for the purpose of classification:

(1) **Tamil Genres**: The database has been formed with clippings from commercially available Tamil music CDs. The classical database is from devotional songs composed in Carnatic style by Vedanayagam Sastriar, Subramanya Bharathiar and other devotional singers, and the Folk database is formed from Folk songs by popular folk singers like Dr. Pushpavanam Kanda swamy and others.

- Tamil Classical style devotional music – 103 song excerpts,
- Tamil Folk music – 113 song excerpts.

The training set for both genres comprised of 70 songs each, whereas the testing data had 43 Folk songs and 33 classical songs.

(2) **Western Genres**: The dataset contains 10 genres, each represented by 100 tracks, which are each of 30-second duration from the GTZAN database that is available online. The tracks are all 22 050 Hz Mono 16-bit audio files in .wav format. 100 songs each of Rock and Classical have been chosen for the purpose of classification. Rock music has a fast rhythm and beat like the Tamil Folk music. The number of songs for the training set and the test set for the Rock and Classical genres were chosen to be 60 and 40 respectively.

Since the chorus of a song is more descriptive of the genre, the 30-second excerpts were taken from the middle of each song, approximately 2 minutes after the beginning of each piece for both the Tamil and western genres.

3. Experimental results

The results obtained from experimentation have been discussed in this section. The 30-second song excerpts of both Tamil and western genres were framed into 20 ms frames with a 50% overlap so that one feature has been obtained every 10 ms. Timbral, Spectral Shape, Temporal, MFCC, Fractional MFCC and Statistical features were extracted from the excerpts.The statistical values of the Mean and Standard deviation were calculated from the temporal summarization of the feature values, along the 20 ms frames. The features were sorted to identify the best feature descriptors for the Tamil and western genres. The Mean and Standard deviation of the features were fed as feature input vectors to the classifiers. K-NN and SVM classifiers were used for Tamil genres whereas only SVM was employed for the western genres. The results were observed as follows:

(1) **Tamil genres:** The features that contributed for efficient discrimination of Tamil music were Spectral Roll-off, Flux, Skewness, Kurtosis and FrFT based MFCC. The respective graphs of the contributing features have been shown below.

For this experiment the value of $K = 2$ gave a higher accuracy than other values. The Radial Basis Function (RBF) was used as the kernel function for

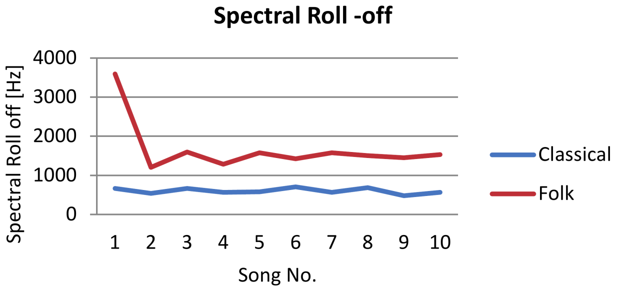

Fig. 5. Spectral Roll-off values of Classical and Folk genres.

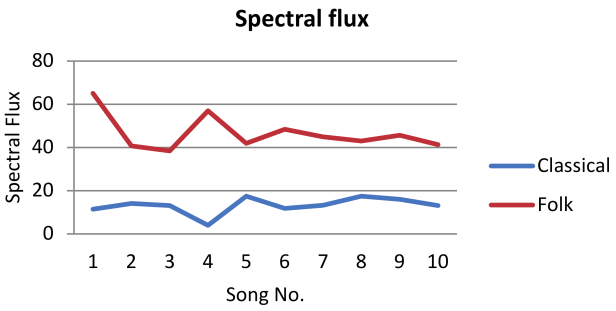

Fig. 6. Spectral Flux values for Classical and Folk genres.

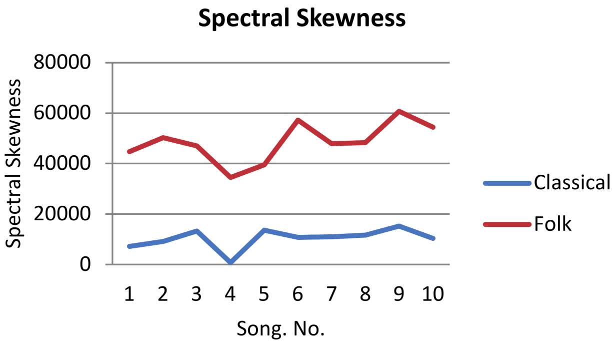

Fig. 7. Spectral Skewness for Classical and Folk genres.

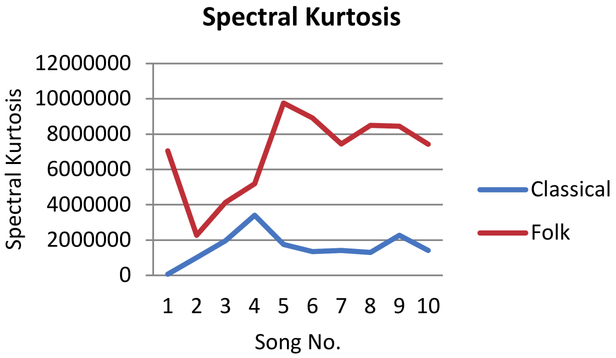

Fig. 8. Spectral Kurtosis values for Classical and Folk genres.

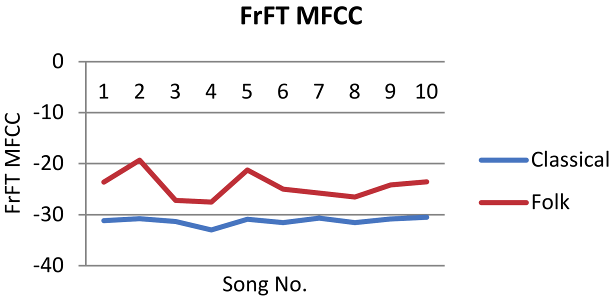

Fig. 9. Fractional MFCC values for Classical and Folk genres.

the SVM classifier, since it efficiently handles multi-class problems. The value for gamma of the RBF kernel was chosen to be 0.5 and the cost function $C = 5$, for the purpose of this experiment. The results have been displayed in Table 2.

Table 2. Accuracies for different combinations of features with KNN and SVM classifiers for Tamil genres.

Feature Set	Classifier	% Accuracy
Feature Set 1: Spectral Roll off + Flux + Skewness + Kurtosis + MFCC	KNN SVM (RBF Kernel)	66.23 83.50
Feature Set 2: Spectral Roll off + Flux + Skewness + Kurtosis + Fractional MFCC	KNN SVM (RBF Kernel)	70.45 96.05

Table 3. Confusion matrix for features with MFCC (Tamil genres) and FrMFCC (Tamil genres).

Number of songs = 76		Predicted		
		Classical	Folk	
MFCC (Tamil genres)				
Actual	Classical	32 (TP)	1 (FN)	33
	Folk	2 (FP)	41 (TN)	43
		34	**42**	
FrMFCC (Tamil genres)				
Actual	Classical	21 (TP)	12 (FN)	33
	Folk	0 (FP)	43 (TN)	43
		21	**55**	

TP – True Positive gives the number of correct predictions that Classical song is "Classical",

FN – False Negative is the number of incorrect predictions that a Classical song is "Folk",

FP – False Positive is the number of incorrect predictions that a Folk song is "Classical",

TN – True Negative number of correct predictions that a Folk song is "Folk".

Table 4. Performance Measures for Feature Set 1 and Feature Set 2 (Tamil genres).

Performance measure	Feature Set 1 (with MFCC) [%]	Feature Set 2 (with FrMFCC) [%]
Accuracy (TP+TN)/ (TP+TN+FP+FN)	84.21	96.05
Precision (TN/(FN+TN))	78.00	97.61
Recall (TN/(FP+TN))	95.34	100.00

Accuracy: The proportion of the total number of correct predictions.

Precision: The proportion of the predicted positive cases that were correct.

Recall: The proportion of positive cases that were correctly identified.

(2) **Western Genres:** From the experimentation, it was observed that the FrFT based MFCC when combined with other spectral features significantly increased the accuracy of western genres as well. But it was also observed that the feature combination for accurate classification of western genres was different from that of the Tamil genres. While spectral skewness and kurtosis greatly contributed to improve the classification accuracy of Tamil genres, these two features did not make any contribution to the classification of the two western genres of rock and classical music. The two features along with spectral centroid infact reduced the classification accuracy. The feature combination of spectral roll-off, spectral flux and Fractional MFCC yielded the highest accuracy of 96.25% compared to an accuracy of 80% for the same features with MFCC. With FrFT based MFCC, 77 songs out of 80 songs were classified correctly. The misclassifications were due to 3 Classical songs being misclassified as Rock.

Table 5. Accuracies for Western Genres of Rock and Classical.

Feature Set	Classifier	% Accuracy
Spectral Centroid + Roll off + Flux + Skewness + Kurtosis + FrMFCC	SVM (RBF kernel)	75.00
Feature Set1: Spectral Roll off + Flux + MFCC		80.00
Feature Set2: Spectral Roll off + Flux + Fractional MFCC		96.25

Table 6. Confusion matrix for Feature set with MFCC (western genres) and FrFT based MFCC (western genres).

Number of songs = 80		Predicted		
		Classical	Rock	
MFCC (western genres)				
Actual	Classical	40 (TP)	14 (FN)	54
	Rock	0 (FP)	40 (TN)	40
		40	**54**	
FrFT based MFCC (western genres)				
Actual	Classical	40 (TP)	3 (FN)	43
	Rock	0 (FP)	40 (TN)	40
		40	**43**	

TP – True Positive gives the number of correct predictions that Classical song is "Classical",

FN – False Negative is the number of incorrect predictions that a Classical song is "Rock",

FP – False Positive is the number of incorrect predictions that a Rock song is "Classical",

TN – True Negative number of correct predictions that a Rock song is "Rock".

Table 7. Performance Measures for Feature Set 1 and Feature Set 2 (western genres).

Performance measure	Feature Set 1 (with MFCC) [%]	Feature Set 2 (with FrMFCC) [%]
Accuracy (TP+TN)/ (TP+TN+FP+FN)	85.10	96.38
Precision (TN/(FN+TN))	74.07	93.02
Recall (TN/(FP+TN))	100.00	100.00

Accuracy: The proportion of the total number of correct predictions.

Precision: The proportion of the predicted positive cases that were correct.

Recall: The proportion of positive cases that were correctly identified.

4. Conclusion and future work

In this paper, a novel feature extraction scheme for automatic genre classification of Indian Tamil music and western music, using combination of FrFT based Fractional MFCC features with Timbral features have been proposed. Since, Fractional Fourier Transform makes use of chirp decomposition that is highly suitable for music signal processing, the proposed FrFT based MFCC features efficiently classifies the Tamil genres and western genres with higher accuracy compared to the conventional MFCC features. For Tamil music, the feature combination of Spectral Roll off, Spectral Flux, Spectral Skewness and Spectral Kurtosis, when combined with Fractional MFCC features, outperforms all other feature combinations, to yield a classification accuracy of 96.05% with an SVM (RBF kernel) classifier. It has also been observed that the FrFT based MFCC along with Spectral Roll-off and Spectral Flux efficiently classifies the western genres (Rock and Classical) from the GTZAN dataset with a higher classification accuracy of 96.25% as compared to the classification accuracy of 80% with MFCC

Acknowledgments

We wish to thank JSPM's Rajarshi Shahu College of Engineering, Savitribai Phule Pune University, Pune, India for providing the lab facilities and guidance.

References

1. Ashok Narayanan V., Prabhu K.M.M. (2003), *The Fractional Fourier Transform: theory, implementationand error analysis*, Microprocessors and Microsystems, **27**, 10, 511–521, doi: 10.1016/S0141-9331(03)00113-3.

2. Bagul M., Soni D., Saravana Kumar K. (2014), *Recognition of similar patterns in popular Hindi Jazz songs by music data mining*, International Conference on Contemporary Computing and Informatics (IC3I), pp. 1274–1278, November 27–29, doi: 10.1109/IC3I.2014.7019799.

3. Baniya B.K., Ghimire D., Lee J. (2014), *A novel approach of automatic music genre classification based on timbral texture and rhythmic content features*, 16th International Conference on Advanced Communication Technology (ICACT), pp. 96–102.

4. Benetos E., Kotropoulos C. (2010), *Non-Negative Tensor Factorization Applied to Music Genre Classification*, IEEE Transactions on Audio, Speech, and Language Processing, **18**, 8, 1955–1967.

5. Bhalke D.G, Rao C.B.R., Bormane D.S. (2014), *Musical Instrument classification using higher order Spectra*, International Conference on Signal Processing and Integrated Networks (SPIN), pp. 40–45, February 20–21, doi: 10.1109/SPIN.2014.6776918.

6. Bhalke D.G., Rao C.B.R., Bormane D.S. (2016), *Automatic musical instrument classification using Fractional Fourier Transform based-MFCC features and counter propagation neural network*, Journal of Intelligent Information System, **46**, 3, 425–446, doi: 10.1007/s10844-015-0360-9.

7. Chen S-H., Chen S-H., Truong T-K. (2012), *Automatic music genre classification based on wavelet package transform and best basis algorithm*, IEEE International Symposium on Circuits and Systems (ISCAS), pp. 3202–3205, May 20–23.

8. Chen S-H., Chen S-H., Guido R.C. (2010), *Music genre classification algorithm based on dynamic frame analysis and support vector machine*, IEEE International Symposium on Multimedia (ISM), pp. 357–361, December 13–15.

9. Fu Z., Lu G., Ting K.M., Zhang D. (2011), *A Survey of audio-based music classification and annotation*, Multimedia IEEE Transactions, **13**, 2, 303–319.

10. Gaikwad S., Chitre A.V., Dandawate Y.H. (2014), *Classification of Indian classical instruments using spectral and principal component analysis based cepstrum features*, International Conference on Electronic Systems, Signal Processing and Computing Technologies (ICESC), pp. 276–279, January 9–11.

11. Ghosal A., Chakraborty R., Chandra Dhara B., Saha S.K. (2012), *Music classification based on MFCC variants and amplitude variation pattern: a hierarchical approach*, International Journal of Signal Processing, Image Processing and Pattern Recognition, **5**, 1, 131–150.

12. Jothilakshmi S., Kathiresan N. (2012), *Automatic music genre classification for Indian Music*, International Conference on Software and Computer Applications (ICSCA 2012), IPCSIT, Vol. 41, pp. 55–59, IACSIT Press, Singapore.

13. Kini S., Gulati S., Rao P. (2011), *Automatic genre classification of North Indian devotional music*, National Conference on Communications (NCC), pp. 1–5, January 28–30, doi: 10.1109/NCC.2011.5734697.

14. Krishnaswamy A. (2003), *Application of pitch tracking to South Indian classical music*, [in:] Proceedings of IEEE International Conference on Acoustics, Speech and Signal (ICASSP '03), Vol. 5, pp. V-557-60, April 6–10, doi: 10.1109/ICASSP.2003.1200030.

15. Kumar V., Pandya H., Jawahar C.V. (2014), *Identifying Ragas in Indian music*, 22nd International Conference on Pattern Recognition (ICPR), pp. 767–772, August 24–28, doi: 10.1109/ICPR.2014.142.

16. Li T., Ogihara M. (2006), *Toward intelligent music information retrieval*, IEEE Transactions on Multimedia, **8**, 3, 564–574.

17. Li T., Tzanetakis G. (2003), *Factors in automatic musical genre classification of audio signals*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 143–146, October 19–22.

18. Lim S-C., Lee J-S., Jang S-J., Lee S-P., Kim M.Y. (2012), *Music-genre classification system based on spectro-temporal features and feature selection*, IEEE Transactions in Consumer Electronics, **58**, 4, 1262–1268.

19. Meng A., Ahrendt P., Larsen J., Hansen L.K. (2007), *Temporal feature integration for music genre classification*, IEEE Transactions in Audio, Speech, and Language Processing, **15**, 5, 1654-1664.

20. Nagavi T.C., Bhajantri N.U. (2011), *Overview of automatic Indian music information recognition, classification and retrieval systems*, International Conference on Recent Trends in Information Systems (ReTIS), pp. 111–116, December 21–23, doi: 10.1109/ReTIS.2011.6146850.

21. Rao P. (2012), *Audio metadata extraction: The case for Hindustani classical music*, International Conference on Signal Processing and Communications (SPCOM), pp. 1–5, July 22–25, doi: 10.1109/SPCOM.2012.6290243.

22. Rosner A., Schuller B., Kostek B. (2014), *Classification of music genre based on music separation into harmonic and drum components*, Archives of Acoustics, **39**, 4, 629–638, doi: 10.2478/aoa-2014-0068.

23. Salamon J., Gome E. (2012), *Melody extraction from polyphonic music signals using pitch contour character-*

istics, IEEE Transactions on Audio, Speech, and Language Processing, **20**, 6, 1759–1770.

24. Scaringella N., Zoia G., Mlynek D. (2006), *Automatic genre classification of music content: a survey*, IEEE Signal Processing Magazine, **23**, 2, 133–141.

25. Shao X., Maddage M.C., Changsheng Xu, Kankanhalli M.S. (2005), *Automatic music summarization based on music structure analysis*, IEEE International Conference on Acoustics, Speech, and Signal Processing, **2**, ii/1169–ii/1172, March 18–23, doi: 10.1109/ICASSP.2005.1415618.

26. Tamil Music, http://www.carnatica.net/tmusic.htm (access on February 2011).

27. Tzanetakis G., Cook P. (2002), *Musical genre classification of audio signals*, IEEE Transactions on Speech and Audio Processing, **10**, 5, 293–302, doi: 10.1109/TSA.2002.800560.

28. Vedanayagam Sastriar, http://www.sastriars.org (access on February 2011).