International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 12, December-2015

Genre Classification of Indian Tamil Music using Mel-Frequency Cepstral Coefficients

Betsy. S
JSPM's Rajarshi Shahu College of Engineering,
Savitribai Phule Pune University,
Pune, India

D. G. Bhalke
JSPM's Rajarshi Shahu College of Engineering
Savitribai Phule Pune University,
Pune, India

***Abstract*—Musical genres are categorical labels created by humans to characterize pieces of music that are related by common characteristics such as instrumentation, rhythmic structure, and harmonic content of the music. This paper presents the Automatic Genre Classification of Indian Tamil Music using Timbral features and Mel-Frequency Cepstral Coefficient features. The classifier model for the proposed system has been built using the K-Nearest Neighbours and Support Vector Machine classifiers. The performance of various features extracted from music excerpts has been analyzed, to identify the appropriate feature descriptors for the two major genres of Indian Tamil music, namely Classical music and Folk music. The results have shown that the feature combination of Spectral Roll off, Spectral Flux, Spectral Skewness and Spectral Kurtosis, combined with Mel-Frequency Cepstral Coefficient features yields a classification accuracy of 84.21 % .**

Keywords—Feature extraction; Mel Frequency Cepstral Coefficients; Timbral features; Tamil music.

NOMENCLATURE

Abbreviation	Description
DCT	Discrete Cosine Transform
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
KNN	K-Nearest Neighbors
MFCC	Mel-Frequency Cepstral Coefficient
MIR	Music Information Retrieval
OSC	Octave-based Specral Contrast
RBF	Radial Basis Function
STFT	Short-Time FourierTransform
SVM	Support Vector Machines
TN	True Negative
TP	True Positive

I. INTRODUCTION

Recent technological advances help users interact with music by directly analyzing the musical content of audio files as consumers have direct online access to thousands of music tracks [17]. The growing digital music databases needs to be properly described and indexed for searching and interaction by end users. MIR deals with the automatic analysis of music signals and uses various characteristics that best describe the music content. Genre information is one such characteristic and is a fundamental component of MIR [18][10].

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. Indian Tamil music can be classified into two main genres, namely 1. Tamil Classical - structured music sung to a rhythmic cycle or tala (Carnatic style music) [19] and 2.Tamil Folk – rural music composed in colloquial style [13].

To characterize any music, various features such as Top-level, mid-level and low-level features have been defined. Top level features are human-defined labels such as genre, mood, and artist. The mid-level features are rhythm features [8] (beat and tempo) and pitch content features (pitch histogram) [5]. The peaks of the autocorrelation function derived from the beat histogram [14] give an idea of the regularity of the music. Short-term or low-level features are timbral features that characterize the spectrum and are derived from a short-time window (or frame). The short-term features commonly used are Spectral centroid/ roll-off/flux, MFCC [15], octave-based specral contrast [12] and time-varying features like zero-crossing and low energy etc. The features extracted from several frames can be integrated to yield temporal features like mean, variance, covariance and kurtosis that give the temporal evolution of the signal across the various frames[2][6].

The features extracted from several frames are integrated to yield temporal features like mean, variance, covariance and kurtosis which will give the temporal evolution of the signal across the various frames.

As much as features are crucial to genre classification, so also is the type of classifier employed. Since the dimensions of features obtained is typically very large, feature selection and dimension reduction techniques like Principal Component Analysis and Linear Discriminant Analysis are employed. Traditional statistical pattern recognition algorithms like KNN, Gaussian Mixture Models, and supervised learning approaches are used to determine the genre. A popular kernel based classifier is the Support Vector Machine, which handle large dimensional feature vectors. Neural Networks and SVM handle large dimensional feature vectors better than other classifiers and can be extended to multiclass classification.

II. RELATED WORK

The pioneering work of George Tzanetakis and Cook (2002) proposing the problem as one of pattern recognition and the free open source software has greatly motivated the research that followed in music genre classification [16]. Tzanetakis and Cook made use of statistical features derived from rhythmic, pitch and timbral content extracted from 30 sec music excerpts and employed classifiers to classify the genres. Tao li and Tzanetakis (2003) extracted features from Daubechies wavelet coefficients histogram and observed that timbral features combined with MFCC yield high accuracy[14]. Xi Shao, Maddage, M.C., Changsheng Xu and Kankanhalli, M.S. (2005) used Linear prediction cepstrum coefficients, zero- crossing rates, MFCC, spectral power, amplitude envelope, spectrum flux, and cepstrum flux with Support vector machines for classifying music [17]. Shin-Cheol Lim Jong-Seol Lee, Sei-Jin Jang, Soek-Pil Lee and Moo Young Kim (2012) worked on classification of genres based on modulation features obtained from Spectro-Temporal features with SVM. The method proved to have higher accuracy at a lower feature dimension for the GTZAN and ISMIR2004 databases [12].

Recent work on Indian music includes the work done by Sujeet Kini, Gulati, S,and Rao, P (2011) on the classification of bhajan and qawwali sub-genres of North Indian devotional music [4]. They have employed timbral features, tempo and modulation spectra of timbral features. By applying 10 fold cross validation of tempo estimations, feature summaries of mean-variance and envelope modulation they obtained an accuracy of 92% with SVM and Gaussian Mixture model. Preethi Rao (2012), worked on the extraction of metadata for Hindustani Classical music using factual information that accompanies music on a CD, such as composer, genre, artist and other semantic labels such as mood [9]. Pitch detection, rhythm detection and melody estimation through motif (repetitive phrase) identification and oscillation (gamakas), were employed for identification of ragas of Hindustani music.

Jothilakshmi .S and Kathiresan.N [3] had worked with various types of Indian music. They had inferred that the feature combination of MFCC, Spectral Centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity feature combination to obtain a classification accuracy of 91.25%, with a Gaussian Mixture Model classifier.

Various features of Timbral, Temporal, Spectral Shape, Wavelet and MFCC have been proposed in the past years. MFCC uses Short Time Fourier Transform (STFT), and has the advantages of linearity, orthogonality and multidimensionality. Thus, it is helps to describe moving signals well. From the above survey, it has been observed that Timbral features and MFCC have made a significant contribution to genre classification [1]. Furthermore, it is observed that no work has been carried out in classifying Indian Tamil music in the past years. Thus, this paper uses Timbral features and MFCC features for the purpose of the classification of Tamil music.

III. FEATURE SET

Features extracted from audio signals give meaningful information, as the audio is characterized by a compact numerical representation. The features extracted are as follows:

A. Timbral Features

The timbral features are obtained from the frequency domain of a signal. The signal is first transformed to the frequency domain and from the spectrum various spectral features are extracted.

Spectral Centroid: The spectral centroid is commonly associated with the measure of the brightness of a sound. This measure is obtained by evaluating the "center of gravity" using the Fourier transforms' frequency and magnitude information. The individual centroid of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes and is given by

$$Spectral\ Centroid = \frac{\sum_{k=1}^{N} kM[k]}{\sum_{k=1}^{N} M[k]} \quad (1)$$

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 roll off is a measure of spectral shape. It is defined as the frequency bin M below which the 85% of the magnitude distribution is concentrated.

$$\text{ɾ}oll\ off = \sum_{k=1}^{M} M[k] \quad (2)$$

where $M[k]$ is the magnitude of the FFT at frequency bin k and N is the number of frequency bins.

Spectral Flux: Spectral Flux reflects the rate of change of power spectrum. It is a measure of how quickly the power changes from frame to frame. It is calculated by comparing the power spectrum of one frame with the power spectrum of the previous frame and is given by

$$Flux = F\left\|M[k] - M_p[k]\right\| \quad (3)$$

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

B. Mel-frequency Cepstral coefficients (MFCC):

Mel-frequency Cepstral Coefficients are extracted based on the Mel scale band pass filters to model the human auditory system. Here the frequency bands are positioned logarithmically (on the Mel-scale) which approximates the human auditory system's response more closely than the linearly-spaced frequency bands. The Fourier Transform is replaced by a Discrete Cosine Transform which uses only

real numbers and has a strong "energy compaction" property; thus, most of the signal information tends to be concentrated in a few low-frequency components of the DCT, which is why only the first 13 components are returned.

The Squared magnitude of 1024 point FFT is calculated to give the power spectrum. The power spectrum is integrated using overlapping triangular filters that are equally spaced in a Mel- frequency scale. The Mel-frequency is related to the linear frequency *f* by the formula,

$$Mel(f) = 1125 * \ln(1+f/700) \quad (4)$$

The Mel scale cepstral coefficients are obtained by the following equation:

$$C_i(m) = \sum_{n=0}^{M-1} S_i(n) \cos[\pi m(\frac{n-0.5}{M})], 0 \leq m \leq M \quad (5)$$

where, *M* is the number of triangular Band pass filters, $S_i(n)$ is the cepstrum of the log filter bank energies and $C_i(m)$ is the MFCC coefficient of the i^{th} frame.

D. Statistical Features

Spectral Skewness: 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

$$\mu = \int (x - \mu_1)^3 f(x) dx \quad (6)$$

where μ is the mean of the distribution

Spectral Kurtosis: Kurtosis is defined as the fourth standardized moment and is defined as

$$Kurtosis = \frac{\mu_4}{\sigma_4} \quad (7)$$

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

IV. CLASSIFIER

For the purpose of classification two classifiers namely, KNN and SVM classifier have been chosen.

KNN(K nearest neighbors): K-Nearest Neighbors 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) [10]. KNN is used in statistical estimation and pattern recognition as a nonparametric technique which does not make any assumptions on the underlying data distribution [18]. The KNN is a lazy learning algorithm that does not use the training samples to perform any generalization, and the entire data is used for the testing phase without discarding any. So in KNN algorithm, there is 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. 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 Machine (SVM): 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 so in a way that the samples of the separate categories are divided by a clear gap; that is as wide as possible [18]. 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

$$K(x, x') = \exp(\frac{\|x-x'\|}{2\sigma^2} \quad (8)$$

where $\|x-x'\|^2$ is the squared Euclidean distance between the feature vectors and σ is a free parameter of the Gaussian radial basis function and the parameter gamma is defined by $\gamma=1/2\sigma^2$. The cost function 'C' is a parameter indicating the soft margin that controls the influence of each support vector (vectors that define the hyper plane).

V. DATASET

In this work, the database was formed from song clippings taken from various commercially available Tamil gospel Carnatic [19], patriotic and Folk compositions. 103 song excerpts from classical based devotional hymns and 113 song excerpts from folk songs formed the database. 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.

VI. EXPERIMENTAL RESULTS

The results obtained from experimentation are discussed in this section. The 30-second excerpts from 213 songs. The training and testing data were divided approximately in a 60:40 ratio. 137 songs formed training data, and 76 songs were used for testing. The 30-second excerpts were further framed into 20 millisecond frames with a 50% overlap, so that one feature is obtained every 10 millisecond. Timbral features, Spectral Shape features, Temporal and MFCC features were extracted from the excerpts. The features were sorted to identify the best feature descriptors for the genres of Tamil music. The statistical values of Mean and Standard Deviation derived from the temporal summarization of each feature of the texture windows were calculated and fed as feature inputs to the KNN and SVM classifiers. The following graphs describe the features that sufficiently discriminate the two genres.

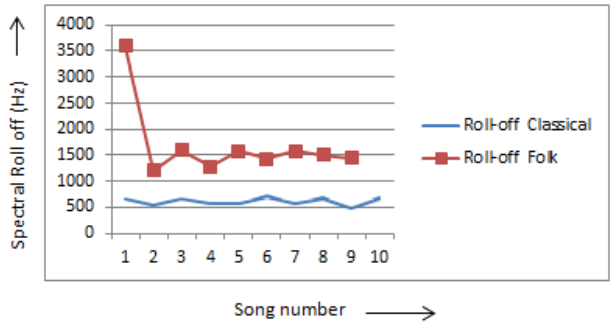

Fig. 1. Roll off values for Classical and Folk genres

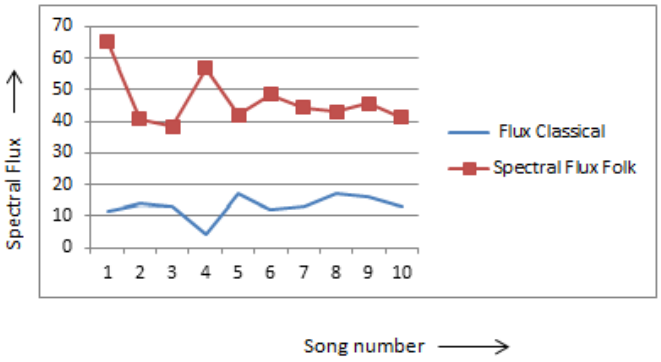

Fig. 2. Spectral flux values for Classical and Folk genres

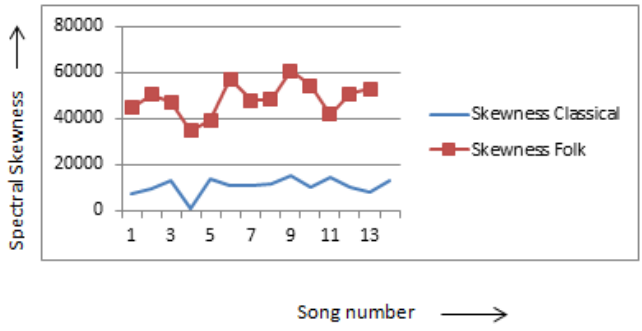

Fig. 3. Spectral Skewness of Classical and Folk genres

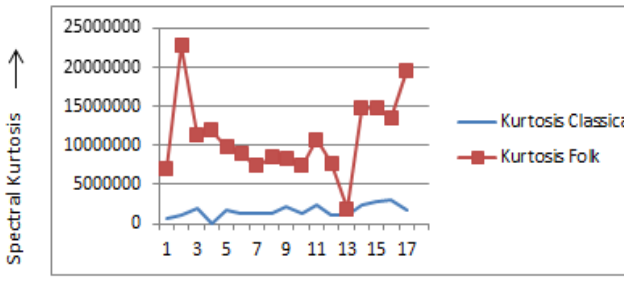

Fig. 4. Spectral Kurtosis of Classical and Folk genres

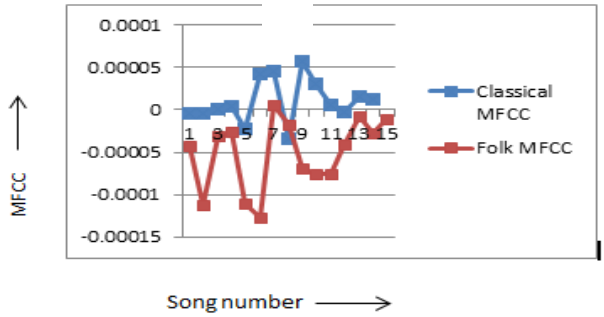

Fig. 5. MFCC values for Classical and Folk genres

The mean and standard deviation of the best feature descriptors namely Spectral Roll-off, Flux, Skewness, Kurtosis and MFCC along the frames of the 30 sec song have been calculated and fed to the KNN and SVM classifiers. For this experiment, 'K'= 2 was used in the KNN, since it was found to yield a higher accuracy. The Radial Basis Function (RBF) was used as the kernel function for the SVM classifier, since it efficiently handles multiclass problems. The value for gamma of the RBF kernel was chosen to be 0.5 and the cost function C =5, was chosen for the purpose of this experiment. The classification results have been displayed in Table 1.

TABLE I. CLASSIFICATION ACCURACIES FOR THE FEATURE COMBINATION WITH KNN AND SVM CLASSIFIERS

Feature Set	*Classifier*	*Accuracy (%)*
Spectral Roll off + Flux + Skewness + Kurtosis + MFCC	*KNN*	66.23
	SVM	84.21

TABLE II. CONFUSION MATRIX FOR THE FEATURE COMBINATION WITH SVM CLASSIFIER

Number of songs=76		Predicted		
		Classical	Folk	
Actual	Classical	21 (TP)	12 (FN)	33
	Folk	0 (FP)	43 (TN)	43
		21	55	

TP- True Positive is the number of predictions that Classical song is "Classical".
FN- False Negative is the number of predictions that a Classical song is "Folk".
FP- False Positive is the number of predictions that a Folk song is "Classical".
TN- True Negative is the number of predictions that a Folk song is "Folk".

The above results show that the number of classical songs misclassified as folk is higher. This may be due to the fast rhythms and tempo of some classical songs. It can also be noted that all folk songs have been correctly classified. The drop in accuracy is fully due to classical songs being misclassified as folk.

TABLE III. PERFORMANCE MEASURES WITH SVM CLASSIFIER

Performance Measure	FEATURE SET (with MFCC)(%)
Accuracy (TP+TN)/(TP+TN+FP+FN)	84.21
Precision (TN/(FN+TN))	78
Recall ((TN/(FP+TN))	100

VII. CONCLUSION AND FUTURE SCOPE

In this paper, a novel feature extraction scheme for automatic genre classification of Indian Tamil music, using MFCC and Timbral features has been proposed. Since, MFCC has the advantages of linearity, orthogonality and multidimensionality that is highly suitable for music signal processing, it gives a good accuracy in discriminating music genres. Moreover, the feature combination of Spectral Roll off, Spectral Flux, Spectral Skewness and Spectral Kurtosis, when combined with MFCC features, outperforms all other feature combinations, to yield a classification accuracy of 84.21 % with an SVM (RBF kernel) classifier.

As Tamil music has many sub-genres like light classical, urban folk music and Thalatu (Lullaby songs), the future work will include investigations on the sub-genres of Classical and Folk music and also on the addition of new features that will improve the accuracy.

ACKNOWLEDGMENT

The author wishes to thank JSPM's Rajarshi Shahu College of Enginneering, Savitribai Phule Pune University, Pune, India for providing the lab facilities.

REFERENCES

[1] Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara and Sanjoy Kumar Saha, "Music Classification based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach ", in *International Journal of Signal Processing*, Image Processing and Pattern Recognition Vol. 5, No. 1, March, 2012

[2] Baniya, B.K.; Ghimire, D.; Joonwhoan Lee, "A novel approach of automatic music genre classification based on timbral texture and rhythmic content features," in *16th International Conference on Advanced Communication Technology (ICACT)*, 2014, vol., no., pp.96,102, 16-19 Feb. 2014

[3] Jothilakshmi. S., Kathiresan . N. "Automatic Music Genre Classification for Indian Music", 2012 International Conference on Software and Computer Applications (ICSCA 2012) , IPCSIT vol. 41 (2012) © (2012) IACSIT Press, Singapore

[4] Kini, S.; Gulati, S.; Rao, P., "Automatic genre classification of North Indian devotional music," National Conference on Communications (NCC), 2011, pp.1-5, 28-30 Jan. 2011doi: 10.1109/NCC.2011.5734697

[5] Krishnaswamy, A., "Application of pitch tracking to South Indian classical music," in the Proceedings of IEEE International Conference on Acoustics, Speech and Signal, 2003. Proceedings. (ICASSP '03), vol.5, no., pp.V-557-60 vol.5, 6-10 April 2003 doi: 10.1109/ICASSP.2003.1200030

[6] Meng, A.; Ahrendt, P.; Larsen, J.; Hansen, L.K., "Temporal Feature Integration for Music Genre Classification", IEEE Transactions in Audio, Speech, and Language Processing, vol.15, no.5, pp.1654-1664, July 2007

[7] Nagavi, T.C.; Bhajantri, N.U., "Overview of automatic Indian music information recognition, classification and retrieval systems", International Conference on Recent Trends in Information Systems (ReTIS), 2011, pp.111-116, 21-23 Dec. 2011 doi: 10.1109/ReTIS.2011.6146850

[8] ROSNER, Aldona; SCHULLER, Bjorn; KOSTEK, Bozena, "Classification of Music Genre Based on Music Separation into Harmonic and Drum Components", Archives of Acoustics [S.I] v.39, n.4, p.629-638, dec 2014 ISSN 2300-262X doi :10.2478/aoa-2014-0068

[9] Rao, P., "Audio metadata extraction: The case for Hindustani classical music", International Conference on Signal Processing and Communications (SPCOM), 2012, vol., no., pp.1-5, 22-25 July 2012 doi: 10.1109/SPCOM.2012.6290243

[10] Scaringella, N.; Zoia, G.; Mlynek, D., "Automatic genre classification of music content: a survey," IEEE Signal Processing Magazine, vol.23, no.2, pp.133,141, March 2006

[11] Shih-Hao Chen; Shi-Huang Chen; Trieu-Kien Truong, "Automatic music genre classification based on wavelet package transform and best basis algorithm," IEEE International Symposium on Circuits and Systems (ISCAS), 2012, pp.3202-3205, 20-23 May 2012

[12] Shin-Cheol Lim; Jong-Seol Lee; Sei-Jin Jang; Soek-Pil Lee; Moo Young Kim, "Music-genre classification system based on spectro-temporal features and feature selection," IEEE Transactions in Consumer Electronics, vol.58, no.4, pp.1262-1268, November 2012

[13] Tamil Music http://www.carnatica.net/tmusic.htm

[14] Tao Li; Tzanetakis, G., "Factors in automatic musical genre classification of audio signals," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2003, pp.143-146, 19-22 Oct. 2003

[15] Tao Li, Ogihara, M., "Toward intelligent music information retrieval", IEEE Transactions on Multimedia, vol: 8, pp: 564-574, June 2006

[16] Tzanetakis, G.; Cook, P., "Musical genre classification of audio signals, "IEEE Transactions on Speech and Audio Processing, vol.10, no.5, pp.293-302, Jul 2002 doi: 10.1109/TSA.2002.800560

[17] Xi Shao; Maddage, M.C.; Changsheng Xu; Kankanhalli, M.S., "Automatic music summarization based on music structure analysis," IEEE International Conference on Acoustics, Speech, and Signal Processing,(2005), vol.2, no., pp.ii/1169-ii/1172 Vol. 2, 18-23 March 2005 doi: 10.1109/ICASSP.2005.1415618

[18] Zhouyu Fu; Guojun Lu; Kai Ming Ting; Dengsheng Zhang, "A Survey of Audio -Based Music Classification and Annotation," Multimedia IEEE Transactions, vol.13, no.2, pp.303, 319, April 2011

[19] VedanayagamSastriar http://www.sastriars.org/