Using MFCC Features for the Classification of Monophonic Music

H.L.Shashirekha
Department of Computer Science
Mangalore University
Mangalore

ABSTRACT
The increase in the availability of music over internet has attracted researchers to devise automated tools and techniques for the classification and retrieval of music in an effective manner. In this paper, we propose an approach to automatically classify the monophonic songs or cappella. Each song in the training set is divided into frames and from each frame thirteen MFCC features are extracted. The average and variance of these features are used to represent each song for two different classifiers. These features are used to train classifiers separately which can then assign a suitable class to an unlabeled song. Experiments are conducted on two different datasets to illustrate the effectiveness of the proposed method.

General Terms
Information retrieval and data mining, Artificial Intelligence Knowledge Management, Speech Processing.

Keywords
Monophonic music, MFCC, Classification.

1. INTRODUCTION
Music Information Retrieval (MIR), a relatively new field of research addresses various issues related to searching and retrieving the digital music content efficiently and effectively from ever increasing music content. The increase in the availability of music over internet in digital form has attracted researchers to devise automated tools and techniques for the classification and retrieval of music in an effective manner. Music classification is the process of automatically assigning a particular music (audio files of sounds, instruments, songs etc.,) to a class or category (singers, instruments, genre etc.,) based on its characteristics (MFCC, chroma, centroid, melody etc.,) [4, 7, 8, 9, 10]. Music in general is polyphonic in the sense that two or more audio sources like voice and playing different instruments are recorded simultaneously. Monophonic music on the other hand also termed as cappella is a simplest form of music from a single source like only voice or melody without any accomplishments. Classifying polyphonic music is more challenging than classifying monophonic music due to the complexity of the problem.

Literature review reveals that most of the research work related to music caters to the needs of western music and very few attempts are reported related to Indian music. Indian music forms which are melodic in nature have certain intricate nuances like raga, tala, thata, pakad, sanchara, gamaka which are not present in western music. As a result MIR of Indian music has got its own share of challenges in addition to the general problems related to processing music. In this paper, we propose an approach to automatically classify the monophonic songs or cappella sung in Kannada, the local language of Karnataka, based on Mel-frequency cepstral coefficients (MFCC).

The remaining part of the paper is organized as follows: Section 2 gives a brief summary of the work done related to music classification. Our methodology and data set creation is discussed in section 3 and experimental results obtained for the classification of monophonic music are presented in section 4. Finally conclusions are drawn and some future work is proposed in section 5.

2. RELATED WORK
Music classification, which is still in its infancy, has received considerable attention of researchers around the globe. As a result, there have been many studies on automatic classification of music according to several criteria using several features and techniques. Several researchers have explored Mel Frequency Cepstral Coefficients (MFCC) for various applications related to audio processing in general and music in particular. Beth Logan [1] have used Mel Frequency Cepstral Coefficients as dominant features for speech recognition and have investigated their applicability in modeling music. MFCC have been used widely from voice recognition to acoustic sound analysis by Z. Jun et.al., [2]. B Sung et. Al., [3] have used simplified version of MFCC and extracted feature data that have audible characteristics from music content, due to which they were able to identify consistent features even if two musical contents used different digitizing specification. A. Kamelia et. al.,[5] have used MFCC for speaker identification by introducing L1-SVM classifier for music genre classification which integrates structural risk minimization benefits of the SVM and the over-fitting resilience of the L1-regression method. A. Ghosal et. al., [8] relies mainly on MFCC for classifying music data hierarchically. At the top level music is classified as song (music with voice) and instrumental (music without voice) and these in turn are classified based on instrument type and genres respectively. L. Yongchun et. al., [9] proposes a music classification system based on Back Propogation Neural Network (BPNN) where in the features of the music data is MFCC. Loughran, R. et. al., [11] makes use of MFCC for the classification of musical instruments. They analyze the samples of few instruments to get these coefficients and these coefficients are reduced using principal component analysis. F. Gouyon et. al., [13] have considered MFCC like descriptors as a specific set of rhythmic descriptors for which they provide procedures of automatic extraction from audio signals. MFCC is used for speaker identification by Md. Rashidul Hasan et. al., [6], where speaker voice is used to verify the identity through which they can control access to services such as voice dialing, banking by telephone etc. G. Agostini et al., [4] focuses on the recognition of musical instruments out of monophonic musical signals, aiming to achieve a compact representation. Michael I. M. and D.P.W. Elli [7] defines a system for identifying artist, which uses support vector machines to classify songs based on features calculated over their entire lengths. J. Shih et. al., [10] have proposed two novel music features namely low-frequency

energy ratio and energy domain signal coding to facilitate automatic music genre classification. A unifying framework for feature extraction is proposed by I. Mierswa and K. Morik [12], for classification of genres and classification according to user preferences. Chai, W, and B. Vercoe [14] describe their work on the classification of folk music from different countries based on their monophonic melodies using hidden Markov models. F. Fernandez et. al., [15] have presented a preliminary attempt to apply Fuzzy Rule-Based System in cooperation with Evolutionary Algorithms to musical genre classification. In spite of several attempts by researchers for various applications there are many more challenges related to music processing which need to be addressed.

3. PROPOSED WORK

The major steps in classification are preprocessing, feature extraction, classifier model construction and model evaluation. As the music recordings were done in an isolated room, for simplicity it is assumed that the music recordings considered as data in this work do not contain any major noise and hence preprocessing that includes noise elimination is skipped. The frame work used in this work is as shown in figure 1.

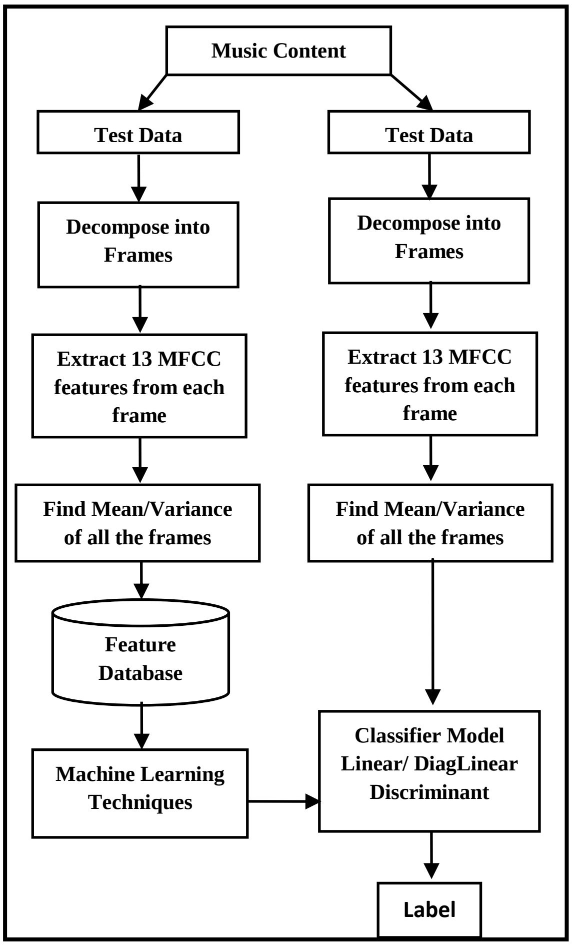

Fig 1: Frame work of our approach

3.1 Feature Extraction

Feature extraction plays a major role in deciding the performance of a classifier. Features such as pitch or melody, tempo, genre, rhythm, timbre, spectral, can be extracted from music files depending upon the application. MFCCs are short term spectral based features [1] used by many researchers for speech recognition, music retrieval system, music summarization, speech/music discrimination. The strength of MFCC lies in its ability for compact representation of amplitude spectrum. MFCCs have been the most widely used features for speech recognition due to their ability to represent the speech spectrum in a compact form. MIRtoolbox [16], an integrated set of functions written in Matlab, dedicated to the extraction of musical features from audio files is used in this work. The features are extracted by analyzing the recordings using MIRtoolbox as explained below:

A song or a music signal in the training set is divided into frames where each frame is of 0.04 seconds with 50% overlap. From each frame 13 MFCC features are extracted and then the mean and variance of the MFCC's of all the frames are computed separately for each song. The mean and variance of the MFCC's are then used to train the classifier which can be used to assign a suitable class label to the unlabeled music signal or song.

3.2 Classifier Model Construction

Linear discriminant function and diaglinear discriminant function are the two classifiers used for multiclass categorization of unlabeled music data. These two classifiers are available as default classifiers in Matlab. Discriminant functions are the methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or more commonly for dimensionality reduction before classification. While Linear discriminant function fits a multivariate normal density to each group with a pooled estimate of covariance, diaglinear discriminant function is similar to linear discriminant function but with a diagonal covariance matrix estimate. The features extracted from the music data is used to construct the classifiers.

3.3 Data Set Creation

As there are no standard datasets for Indian music to test the performance of our approach, an attempt is made to create a small dataset of Kannada songs. This dataset includes 65 songs without any accomplishments belonging to 11 different songs representing 11 different melodies or classes. Each song in a class is sung by 5 to 8 singers whose musical experience ranged from naïve to professional. The singers were aware that the recordings will be used as dataset in an experiment on MIR and the songs were recorded in an uncontrolled but no noisy environment. There was no restriction as to how much of the song should be sung. As a result some recordings include complete songs, while some include only a significant portion of the song that covers the melody of that song. All recordings were done using a basic laptop microphone and no post-processing was applied. These 65 audio files of were recorded as mono WAV format sampled at 44.1kHz.

4. EXPERIMENTS AND RESULTS

All implementations were carried out in Matlab8.0 on Windows 7 platform. The experiments were performed using two classifiers on two different data sets. The algorithms were tested on the newly created music collection of Kannada songs and on a subset of GTZAN Genre Collection [17]. The Genre dataset consists of 1000 audio tracks each of 30 seconds long. It contains 10 genres namely Blues, Classical, country, Disco, Hiphop, Jazz, Metal, Pop, Reggae and Rock and there are 100 tracks for each genre. These genres are labeled from 1 to 10 respectively. The tracks are all 22050Hz Mono 16-bit audio files in .wav format. Instead of the entire Genre collection we have considered 10 genres, each having 50 tracks.

4.1 Results

4.1.1 Kannada Songs Dataset

The experiment is repeated 5 times each time changing the training set and test set and computing the accuracy of the classifier. Accuracy obtained in all the iterations are averaged and displayed in the form of confusion matrix. The confusion matrix for linear classifier using mean and variance of MFCC features is shown in Table-1 and Table-2 respectively. Figure-2 gives the comparison of the performances of the linear classifier using mean and variance of MFCC features in all the 5 iterations. Similarly, confusion matrix for diaglinear classifier using mean and variance of MFCC features is shown in Table-3 and Table-4 respectively. Figure-3 gives the comparison of the performances of the diaglinear classifier using mean and variance of MFCC features.

Table 1. Confusion matrix for linear classifier using mean of MFCC features for Kannada songs dataset

Class Labels	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	1	0	2	1	1	0	0
2	0	5	0	0	0	0	0	0	0	0	0
3	1	0	1	2	0	0	0	1	0	0	0
4	0	1	0	1	2	0	0	0	1	0	0
5	1	0	0	1	0	0	1	1	0	1	0
6	1	1	1	0	0	0	0	1	1	0	0
7	1	0	0	0	0	1	1	0	2	0	0
8	0	0	2	0	1	0	0	0	1	0	1
9	1	0	1	0	1	0	1	0	1	0	0
10	0	0	0	0	0	0	0	1	0	4	0
11	0	0	0	1	0	0	0	0	0	1	3

Table 2. Confusion matrix for linear classifier using variance of MFCC features for Kannada songs dataset

Class Labels	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	1	0	1	0	2	0	0	1
2	0	2	0	0	0	1	1	0	0	1	0
3	1	1	1	0	0	0	0	1	0	1	0
4	0	1	0	2	0	2	0	0	0	0	0
5	0	0	1	0	1	0	1	2	0	0	0
6	0	0	2	0	1	2	0	0	0	0	0
7	1	0	1	0	1	0	0	1	1	0	0
8	0	0	1	0	2	0	0	1	1	0	0
9	0	0	0	2	0	1	0	0	2	0	0
10	0	1	0	0	1	0	1	0	0	2	0
11	0	0	1	0	0	0	1	1	0	0	2

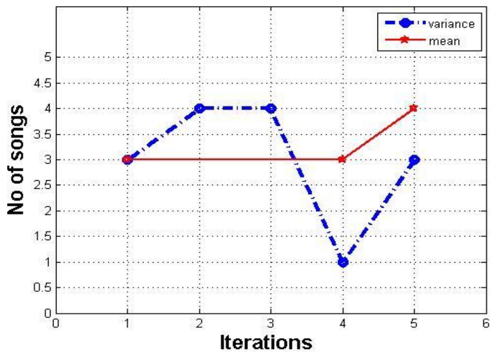

Fig 2: Comparison of the performances of linear classifier for Kannada songs dataset

Table 3. Confusion matrix for diaglinear classifier using mean of MFCC features for Kannada songs dataset

Class Labels	1	2	3	4	5	6	7	8	9	10	11
1	0	0	1	1	0	1	1	1	0	0	0
2	0	5	0	0	0	0	0	0	0	0	0
3	0	0	1	2	0	0	1	1	0	0	0
4	0	1	0	3	1	0	0	0	0	0	0
5	0	0	0	1	0	1	0	1	0	1	1
6	0	0	1	1	1	1	1	0	0	0	0
7	0	0	0	0	0	2	1	0	1	1	0
8	0	0	1	0	1	1	0	0	0	1	1
9	0	0	1	1	0	0	0	1	1	1	0
10	0	0	0	0	0	0	0	0	0	4	1
11	0	0	0	0	0	0	0	0	0	2	3

Table 4. Confusion matrix for diaglinear classifier using variance of MFCC features for Kannada songs dataset

Class Labels	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	2	0	1	0	1	1
2	0	3	1	0	0	0	0	0	0	1	0
3	0	0	0	1	0	2	0	2	0	0	0
4	0	3	0	1	0	1	0	0	0	0	0
5	0	0	0	2	0	1	0	2	0	0	0
6	0	1	0	2	0	2	0	0	0	0	0
7	0	0	0	1	1	1	0	1	0	1	1
8	0	0	0	0	1	1	0	1	0	1	1
9	0	0	0	0	0	2	0	1	0	2	0
10	0	1	0	0	0	0	0	1	0	2	1
11	0	0	0	0	0	0	0	1	0	0	4

The low performance of the classifier may be due to many reasons including the music expertise of the singers, quality of the music, the presence of inherent noise, due to less number of features and also due to the classifiers.

4.1.2 GTZAN Genre Collection

The experiment is repeated 5 times as discussed in section 4.1.1. The confusion matrix for linear classifier using mean and variance of MFCC features is shown in Table-5 and Table-6 respectively. Figure-4 gives the comparison of the performances of the linear classifier using mean and variance of MFCC features in all the 5 iterations. Similarly confusion matrix for diaglinear classifier using mean and variance of MFCC features is shown in Table-7 and Table-8 respectively. Figure-5 gives the comparison of the performances of the diaglinear classifier using mean and variance of MFCC features.

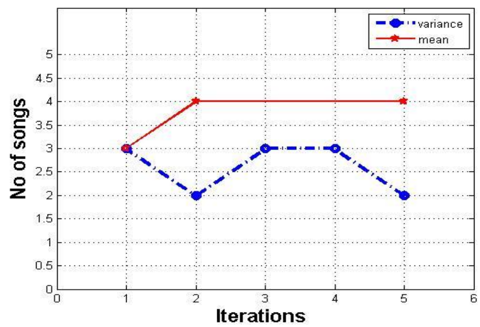

Fig 3: Comparison of the performances of diaglinear classifier for Kannada songs dataset

Table 5. Confusion matrix for linear classifier using mean of MFCC features for GTZAN Genre Collection

Class Labels	1	2	3	4	5	6	7	8	9	10
1	12	2	9	1	1	13	1	4	5	2
2	0	40	1	0	0	8	0	1	0	0
3	4	2	10	4	6	9	1	2	7	5
4	0	1	2	19	9	2	2	8	3	4
5	0	0	3	12	20	1	5	1	5	3
6	4	8	10	0	3	18	1	1	0	5
7	0	0	0	1	4	0	39	3	1	2
8	3	3	1	7	7	0	0	23	2	4
9	2	0	5	5	3	1	1	5	21	7
10	3	1	6	13	3	4	3	2	7	8

Table 6. Confusion matrix for linear classifier using variance of MFCC features for GTZAN Genre Collection

Class Labels	1	2	3	4	5	6	7	8	9	10
1	11	3	4	1	8	3	2	5	7	6
2	2	32	0	0	0	11	4	0	0	1
3	1	2	19	4	5	0	8	6	2	3
4	0	0	10	18	3	1	4	3	1	10
5	4	0	5	4	15	0	5	3	11	3
6	1	12	0	2	2	25	4	1	1	2
7	1	0	1	2	1	0	38	1	0	6
8	5	1	8	8	3	0	2	7	11	5
9	6	0	5	3	12	0	0	5	18	1
10	2	0	9	8	2	0	12	3	2	12

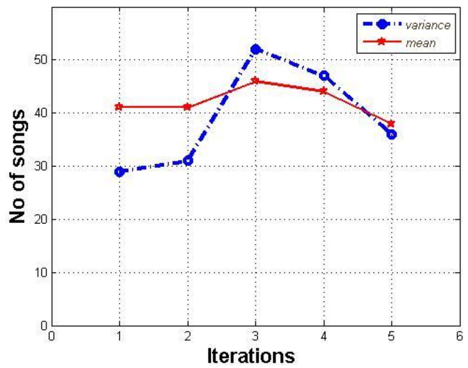

Fig 4: Comparison of the performances of linear classifier for GTZAN Genre Collection

Table 7. Confusion matrix for diaglinear classifier using mean of MFCC features for GTZAN Genre Collection

Class Labels	1	2	3	4	5	6	7	8	9	10
1	16	0	10	0	0	7	0	2	10	5
2	0	42	2	0	0	6	0	0	0	0
3	3	1	11	3	7	9	1	6	5	4
4	0	1	2	17	8	2	1	8	5	6
5	0	0	2	13	17	1	6	2	7	2
6	5	10	10	0	3	15	1	2	1	3
7	0	0	0	0	6	0	40	3	1	0
8	2	3	1	6	5	0	0	24	3	6
9	4	1	2	2	3	1	1	4	24	8
10	3	1	7	12	2	3	3	1	7	11

Table 8. Confusion matrix for diaglinear classifier using variance of MFCC features for GTZAN Genre Collection

Class Labels	1	2	3	4	5	6	7	8	9	10
1	6	3	4	0	7	2	3	5	13	7
2	1	28	0	1	0	11	6	0	0	3
3	2	1	13	6	6	1	3	3	5	10
4	1	0	7	16	4	0	5	3	0	14
5	4	0	2	6	15	1	3	2	13	4
6	1	8	0	3	2	18	11	0	1	6
7	1	0	0	1	1	0	43	1	0	3
8	4	1	7	12	4	1	0	5	9	7
9	6	1	3	4	11	0	0	1	19	5
10	1	0	4	8	2	0	17	1	1	16

5. CONCLUSION AND FUTURE WORK

In this research work, two simple classification approaches based on MFCC features and an attempt to create a dataset of Kannada songs sung by people with varying musical expertise are presented. This is a first step towards processing Indian music in general and processing Kannada songs for various applications in particular. The extension of this work will include extracting some more valuable features and experimenting on a large dataset with different types of classifiers. Even though the methodology presented here reports low performance, it could be improved by considering more features at different levels and also considering large and exhaustive data set.

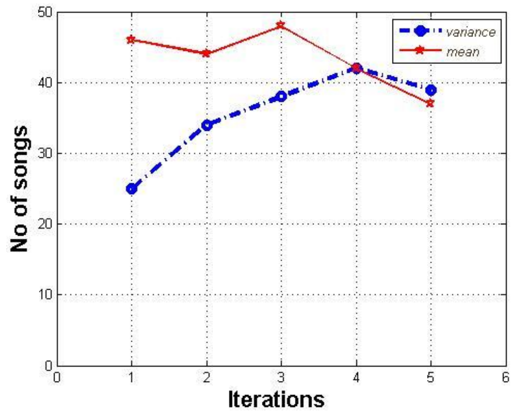

Fig 5: Comparison of the performances of diaglinear classifier for GTZAN Genre Collection

6. ACKNOWLEDGMENTS

I thank Mrs. Rashmi N M, student of IV semester, M.Sc (Computer Science), Mangalore University, for rendering her help in implementing this research work.

7. REFERENCES

[1] Beth Logan, "Mel frequency cepstral coefficients for music modeling", In Proc. of the International Symposium on Music Information Retrieval, 2000.

[2] Z. Jun, S. Kwong, W. Gang, Q.Hong, "Using Mel Frequency Cepstral Coefficients in Missing Data Technique", In EURASIP Journal on Applied Signal Processing, Vol.2004, No. 3, pp. 340-346, 2004.

[3] Bokyung Sung, Myung Bum Jung and Ilju Ko, "A Feature Based Music Content Recognition method using Simplified MFCC" International Journal of Principles

and Applications of Information Science and Technology, Vol.2, No.1, pp. 13-23, July 2008.

[4] G. Agostini, Maurizio Longari, Emanuele Pollastri, “Musical Instrument Timbres Classification with Spectral Features”, In EURASIP Journal on Applied Signal Processing vol 2003, no. 1, pp 5–14, 2003.

[5] Aryafar, Kamelia, Sina Jafarpour, and Ali Shokoufandeh “Music genre classification using sparsity-eager support vector machines”, Technical report, Drexel University, 2012.

[6] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman, “Speaker Identification Using Mel Frequency Cepstral Coefficients”, In the Proceedings of 3rd International Conference on Electrical & Computer Engineering, ICECE 2004, pp. 565-568, December 2004.

[7] Michael I. Mandel and Daniel P.W. Ellis, “Song-Level Features And Support Vector Machines For Music Classification”, In the Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), pp.594-599, September 2005.

[8] 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, pp. 131-150, March, 2012.

[9] LiuYongchun, Song Hong, Yang Jing, “Research on Music Classification Based on MFCC and BP Neural Network”, In the Proceedings of 2nd International Conference on Information, Electronics and Computer, ICIEAC, pp. 129-132, 2014.

[10] Jau-Ling Shih, Chang-Hsing Lee, Shiue-Wei Lin, “Automatic Classification of Musical Audio Signals”, In Journal of Information Technology and Applications, Vol. 1 No. 2, pp. 95-105, September, 2006.

[11] Loughran, R., Walker, J., O'Neill, M., and O'Farrell, M., “The Use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification”, In the Proceedings of International Computer Music Conference (ICMC), 24-29 August 2008

[12] I. Mierswa and K. Morik, “Automatic Feature Extraction for Classifying Audio Data”, In Machine Learning Journal, vol.58, Issue 2-3, pp.127-149, 2005.

[13] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, “Evaluating rhythmic descriptors for musical genre classification”. In the Proceedings of the AES 25th International Conference, London, UK, June 17-19 2004.

[14] Chai, W, and B. Vercoe, “Folk music classification using hidden Markov models”, In the Proceedings of the International Conference on Artificial Intelligence, June 2001.

[15] F. Fernandez, F. Chavez, R. Alcala, and F. Herrera, “Musical genre classification by means of fuzzy rule-based systems: A preliminary approach”, In IEEE Congress on Evolutionary Computing, pp. 2571–2577, 2011.

[16] O. Lartillot and P. Toiviainen. MIR in matlab (II): A toolbox for musical feature extraction from audio. In Proceedings of 5th International Conference on Music Information Retrieval, 2007.

[17] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals ", In IEEE Transactions on Audio and Speech Processing, Vol. 10, Issue 5, pp. 293-302, 2002.