Journal of Research and Applications (JRA), UCSMTLA, 2019 Volume-01, Issue-01

Article ID - JRA190031

A STUDY ON FEATURE ANALYSIS USING SPARSE REPRESENTATION FOR MUSIC CLASSIFICATION

May Thu Myint [(1)], Phyu Phyu Khaing [(2)]

[(1)]University of Computer Studies (Hpa-an), Myanmar

[(2)] Myanmar Institute of Information Technology, Mandalay, Myanmar

[(1)]*mthumyint@gmail.com*

ABSTRACT

This paper presents the first attempt to classify for Myanmar ethnic music by using sparse representation classification method to define the class label according to their ethnic traditional style. In this system, the only five myanmar ethnic groups are considered such as Kachin, Kayin, Mon, Shan, Rakhine. The classification system describe the better accuracy by analysing the temporal features and spectral features, the best outcome by calculating all the results based on Sparse Representation classifier in compared with K Nearest Neighbors classifier. Therefore, this audio classification achieved the results by evaluating feature combination and the best feature combination by using SRC and KNN classfier. With all the features of all ethnic classes, the overall outcome of the SRC 64%, which is better than 54% of the overall KNN accuracy. All features (114) combination give the best results of 82% for Kayin ethnic songs than other ethnic songs.The feature combination of MFCC(std, deltamean) are tested on all of five ethnic classes which is the best classification results of 75.00% accuracy from SRC classifier than the classification results of 51.33% from KNN classifier.

KEYWORDS: *myanmar ethnic music, sparse representation classifier, k nearest neighbors, temporal feature, spectral feature*

1. INTRODUCTION

Culture has a great impact on music in term of creation, performance and interpretation. People with a certain cultural background often like a particular cultural music style. Therefore cultural style information is very helpful for listening, searching and recommending music. It can be see that a piece of cultural style music has similar features, similar attributes such as the tuning system, musical scale and instrumentation. On the basis of this observation, machine learning techniques can be used to classify music signals according to the cultural style of music. Automatic music classification is a high-level task that refers to the process of automatically assigning class labels or genre for various tasks, including, but not limited to categorization, organization and browsing. Most audio classification systems combine two processing stages: feature extraction and classification expressed in Liu and others [1]. In this paper, the audio files contained music of different singer, different ethnic song classes (Kachin, Kayin, Mon, Shan,Yakhine etc.) and several other kinds of data such as sounds produced by others cultural people. With these music collections, it would be possible to classify the correct song of the system. In order to provide high quality results and to discriminate more audio music from the huge amount of music collection in the music recognition tasks, the selected features must reflect basic information about the music. In this proposed system, various signal features are used for this purpose including 114 features (zcr, centroid, skew, kurtosis, bandwidth, MFCC mean, MFCC std, MFCC deltamean, MFCC delta- std) have been used for feature extraction in intro songs experiment. Jothilakshmi et.al [2] stated these studies use several different classification strategies, including multivariate Gaussian models, Gaussian mixture models, self-organizing maps, neural networks, k-nearest neighbor schemes and hidden Markov models. The system can be described as a performance evaluation of the proposed system. In addition, evaluation methods are used in a comparative way to measure whether certain changes

lead to an improvement in system performance. The sparse representation classifier evaluate songs by testing several important parameters of them. Many audio classification problems involve high dimensional, noisy data. Sparse representation by l-norm minimization is robust to noise and even incomplete measurements. Now, there are some methods to be compared to the proposed ethnic music classification system, K- Nearest Neighbors classifier (well-known classifier) has to be selected in compared with the sparse representation classifier.

2. EXPERIMENT

2.1 Experiment apparatus

Classification of music based on cultural style can help for music analysis and used in search and recommendation systems. The human perception of the sounds is a way of creating music label during the identification process and as long as they are heard. Technically, people also rely on the features derived from the sound they hear to identify the sound. Deshmukh.et al [3] initiated there are various methods for music classification to classify the ethnic songs and folk songs. Many different types of audio feature extraction have been proposed of the tasks of folk song classification. The system are firstly preprocessed the audio data, from theses audio samples are extracted the nine features (zcr, centroid, shew, centroid kurtosis bandwidth, MFCC mean, MFCC std, MFCC delta-mean and MFCC delta std) which may be calculated based on the basic samples after preprocessing. All of the features are 114 features in which 1-57 features are mean value and 58-114 features are standard deviation value of nine features. Finally, after feature extraction these extracted features are classified by using SRC in comparing with KNN that is to define each ethnic class label.

2.2 Audio Dataset

The audio dataset contained music of different singer, different ethnic song classes (Kachin, Kayin, Mon, Shan, Yakhine) and several other kinds of data such as sounds produced by others cultural people. The dataset contains 250 songs from the popular ethnic music songs categorized as 50 audio recordings of each ethnic classes respectively. The input audio signal (wav file) is resampled at 44100 hz with 16 bits per sample. These songs are from MRTV radio station. Each music piece lasts about 3 to 5 minutes in length but the input music pieces nearly last 60 seconds segment. In all experiments, the intro pieces of music is used for evaluation of the system.

2.3 Experiment Setup

In all experiment, there are three main components, preprocessing, feature extraction and classification. After converting the input audio from stereo to mono channel and is divided into frames with 100ms, these audio samples frame was taken 50% overlapping frame between the successive frames. The major features of audio sample are extracted from the overlapped frame, and then the extracted features are classified by SRC.The system was implemented by matlab programming language.

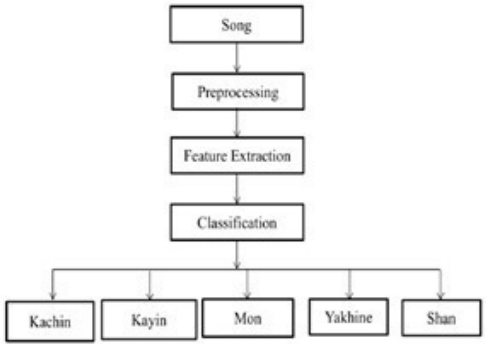

Fig 1. Architecture of myanmar ethnic music classification

From all of the five ethnic classes, in experiment I, the intro pieces of music used for evaluating two feature combination, experiment II ,this ethnic songs dataset is used for analysing three feature combination, also in experiment III for testing four feature combination and the results of experiment IV was achieved by combining five features in comparison SRC with KNN. In experiment V, this experiment is to achieve the best feature combination All songs were trained on the different sets of feature vectors for each consists of nine major features as mentioned in section (2.3).

2.4 Feature Used

The classification system are mainly used spectral and temporal features. At the highest level, music is considered to have four key properties : the melody, or sequence of pitches; the harmony, or the combinations of pitches; the rhythm or organization of sounds in time;; and the timbre or tone color, which is the property that gives each instrument or combination of instruments its distinctive sound. Classification of music would ideally proceed based on these four properties.

2.4.1 Zero Crossing Rate (ZCR)

Acoustic feature ,time-domain feature, and number of times, the signal value crosses zero axis in time domain within a frame. The ZCR also makes it possible to differentiate between voiced and unvoiced speech components: voiced components have much smaller ZCR values than unvoiced ones. The average short-time zero-crossing rate can also be useful in combination with other features in general audio signal classification systems.The ZCR curves are calculated as follows: It is computed as-

$$Z_n = \sum_m |\mathrm{sgn}[x(m)] - \mathrm{sgn}[x(m-1)]| w(n-m),$$

(1)

where sign is the signum function, which returns 1 when the argument is positive, -1 when it is negative, and 0 otherwise.

2.4.2 Mel Frequency Cepstral Coefficients (MFCCs)

The MFCCs stands for the shape of the spectrum with few coefficients. The cepstrum is the Fourier Transform (or Discrete Cosine Transform DCT) of the spectral logarithm. The audio signal is first divided into number of periodic frames. Frames may contain samples that overlap with the previous frame. A window function is also used to minimize deviations at the beginning and end of the frame a windowing function (Hamming window is the most widely used one) is also applied on the frame. The amplitude spectrum for each frame (windowed) is obtained by applying Discrete Fourier Transform (DFT). It has led to the development of Mel frequency. The relation can be expressed as followed in Jothilakshmi et al. [2]:

$$\mathrm{Mel}(f_m) = 2595 \times \log 1 + \frac{f}{700} \qquad (2)$$

2.4.3 Spectral Centroid

It represents the balancing point or the midpoint of the spectral power distribution of a signal. Music involves high frequency sounds which means it will have higher spectral centroid values and the brighter. It is computed as

$$C = \frac{\sum_{f=1}^{N} f * M[f]}{\sum_{f=1}^{N} M[f]}$$

(3)

where the ratio of the sum of spectral magnitude eighted by frequency to the sum of spectral magnitude. A spectral centroid predicts how the dominant frequency of a signal changes over time expressed in Madjarov et al [4].

2.4.4 Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or even undefined described in Bantialebi-Dehkord et.al [5].

- Negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left.
- Positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed, right-tailed, or skewed to the right.

$$skewness = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{ns^3} \qquad (4)$$

2.4.5 Kurtosis

Basically, because kurtosis is a mass movement that does not affect the variance. Consider the case of positive kurtosis, where the heavier tails has a higher peak. In a case of a negative kurtosis, if the mass moves from the tails and center of the distribution to its shoulders,the variance remains the same and the tail is lighter and flatness.

$$kurtosis = \frac{\sum_{i}^{n} (x_i - \bar{x})^4}{ns^4} - 3$$

(5)

where x is the sample mean and s is the sample standard deviation and n is the number of samples. The kurtosis of a normal distribution is 0.

2.5 Classifiers

Many audio classification problems involve high dimensional, noisy data. In this paper, two classifiers are chosen for audio classification as follows:

2.5.1 Sparse Representation Classifier (SRC)

Sparse representation by l_1-norm minimization is robust to noise and even incomplete measurements. For the very large dataset, the SRC classifier is appropriate and thus optimization of the system is crucial. SRC first encodes a query sample into a linear combination of several atoms in a pre-defined dictionary. Then, it identifies the label by assessing which class results in the minimum reconstruction error. The SRC classification methods first calculates the sparse decomposition of test sample on training data set and then calculates the reconstruction residual errors which reconstruct test sample by sparse decomposition coefficients through each class of training samples respectively as in Bantialebi-Dehkord et al. [5]. J.Wright et al. [6] proposed with the SRC, the theoretical step to finding sparse representation is fast if the sparsest solutions are found.. The system finds the optimal description of the music parts from the feature set in respect to more similar function defined in Sparse Representation Classifier (SRC) method.

2.5.2 k-Nearest Neighbors Classifier (*k*NN)

It is a non-linear classifier and the idea is that a small number of neighbors will influence the identification on a point. More precisely, for a specific feature vector in the target set, the *k* closest vectors in the training set are selected and the target feature vector is the label of most representation of k neighbours (there is actually no other training than storing the features of the training set). KNN is the most popular for classification which means the training data is stored so that the classification for unclassified new data will be compared with the training data by taking the data of the most common training. To determine the similarity, the distance function is needed to test the size stated in Jothilakshmi et al. [2].

3. ANALYSIS AND RESULTS

In all experiment, the proposed system is evaluated with the feature combination. There are major nine features in this system. So, the zcr, centroid, bandwidth, MFCC-mean, MFCC-std, MFCC delta-mean, MFCC delta-std are represented by mean value of these features, and also zcr (2), centroid (2), bandwidth (2), MFCC-mean (2), MFCC-std (2), MFCC delta-mean (2), MFCC delta-std (2) are represented by standard deviation value of these features are tested in all experiment. The following classification results are described the feature combination results by combining each feature for each ethnic class .

Experiment I: Evaluation of Two Feature Combination

According to the table 1, fig 2. and fig 3., feature combination of MFCC(std, deltamean) are tested on all of five ethnic classes which is the best classification results of 75.00% accuracy from SRC classifier than the result of 51.33% from KNN classifier. Jothilakshmi et al. [2] observed that performance of KNN is better than GMM for feature combination of MFCC and entropy in the indian music dataset but the low performance is obtained from feature combination of MFCC and spectral centroid for both of KNN and GMM. In our experiment I, the performance of SRC is achieved the best classification accuracy for all two combination of MFCC features. But, the feature combination of MFCC(std,deltastd(2)) give the low classification accuracy of 52.33% from SRC classifier and also 40.33% achieved from KNN classifier for all ethnic songs. In this two feature combination table, all feature combination are achieved the best by using SRC in compared with KNN classifier.

Experiment II: Evaluation of Three Feature Combination

According to table (2), only all feature combination of MFCC are used for myanmar ethnic music dataset which are achieved the better accuracy of 68.33% and 68.30% from SRC than KNN classifier. The performance of SRC is achieved the lowest accuracy of 54% for MFCC(std, mean(2), deltamean(2)) but also KNN classifier gives the low accuracy of 46.67%. In these three feature combination, SRC gives the better accuracy than KNN classifier as shown in fig 4 and fig 5. Jothilakshmi et al. [2] expressed the three feature combination (MFCC, Spectral centroid, Skewness) are used for the indian music dataset in which the performance of KNN and GMM are equally achieved the classification accuracy.

Experiment III: Evaluation of Four Feature Combination

According to table 3 and fig. 6, in the four feature combination of MFCC (mean, std, std(2), deltastd(2)), it obtained the better accuracy of 72.47% and also MFCC (deltamean, deltastd, std(2), deltamean(2)) achieved 70.69% of higher accuracy than the performance of KNN in which 62% and 54.33% for all ethnic songs respectively. In Jothilakshmi et al. [2],

the another combination MFCC, Spectral centroid, skewness, kurtosis, the performance of KNN is achieved 51.25% than 48.13% of GMM. When MFCC(deltamean, deltastd,mean(2),deltastd(2)) are tested with SRC classifier, it achieves the low accuracy of 62.27% but these accuracy is higher than the results of KNN classifier.

Experiment IV: Evaluation of Five Feature Combination

In table (4) and fig 7, the performance of SRC is obtained 70.68% for the feature combination of MFCC (mean, std, deltamean, deltastd, std(2)) that is better than the accuracy of 57% from KNN. When MFCC(mean, std, deltamean, deltastd, deltamean(2)) are combined, the higher accuracy of 66.83% achieved for all ethnic songs than the accuracy of 55.67% from KNN. In Jothilakshmi et al. [2], by adding the flatness feature ,the performance of KNN is better than GMM in indian music classification. According to experiment IV, SRC classifier is achieved the better accuracy for all of MFCC feature combinations than KNN classifier for all ethnic song classes.

Experiment V: Evaluation of Best Feature Combination

According to the table (5), SRC is achieved the best classification result for two feature combination of MFCC (std, delta-mean) and four feature combination of MFCC (mean, std, std(2), deltastd(2)) gives than KNN classifier. Jothilakshmi et.al [2], pointed the performance of KNN is drastically decreased while combining MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity, Rolloff, Spread. Similarly, the results of intro piece of music is descreased by combining all features (114). The performance of SRC is achieved 64% than KNN as shown in the following table 6. In fig 8, SRC gives the best accuracy for Kayin ethnic songs by using all feature combination(114).

4. CONCLUSIONS

In my observation, the system was evaluated by combining two features, three features, four features and five features in many ways. The influence of SRC can behave well these high dimensional audio data compared to other statistical or machine learning methods. Although the classification accuracy is increased to 75% by using the combination of MFCC(std, deltamean) , the classification accuracy of SRC is decreased to 64% when all features (114) are used to classifiy all ethnic classes. From the analysis, the classification system are achieved the best outcomes by combining these timbre features than other features. Therefore, SRC gives the best accuracy of 82% for kayin ethnic songs than other ethnic songs. In conclusion, the obtained results have clearly shown that the system was achieved more representation flexibility and efficiency to classify the ethnic songs. Finally, all of the songs have cultural styles that are played with their respective traditional instruments, in otherwise the classification can't get the better accuracy.

Table 1. Two Feature Combination Classification Accuracy (%) for SRC Vs KNN

Intro Piece of Music (All ethnic classes)					
Feature Combination	**SRC**	**KNN**	**Feature Combination**	**SRC**	**KNN**
MFCC(mean, std)	67.00	54.00	MFCC(std,deltamean(2))	55.67	46.00
MFCC(mean, deltamean)	61.00	55.33	MFCC(std,deltastd(2))	52.33	40.33
MFCCmean, deltastd)	65.60	56.00	MFCC(deltamean,mean(2))	59.00	53.00
MFCC(std,deltamean)	75.00	51.30	MFCC(deltamean,std(2))	69.67	59.00
MFCC(std, deltastd)	59.00	43.33	MFCC(deltamean,deltamean(2))	67.00	59.30
MFCC(mean,mean(2))	60.00	57.00	MFCC(deltamean,deltastd(2))	63.67	51.30
MFCC(mean,std(2))	63.30	52.33	MFCC(deltastd,mean(2))	60.00	53.66
MFCC(mean,deltamean(2))	64.30	57.33	MFCC(deltastd,std(2))	58.66	57.33
MFCC(mean,deltastd(2))	64.66	58.33	MFCC(deltastd,deltamean(2))	61.00	51.33
MFCC(std,mean(2))	57.00	44.66	MFCC(deltastd,deltastd(2))	56.67	48.33
MFCC(std,std(2))	62.34	57.00			

Table 2. Three Feature Combination Classification Accuracy (%) for SRC Vs KNN

Intro Piece of Music (All ethnic classes)					
Feature Combination	**SRC**	**KNN**	**Feature Combination**	**SRC**	**KNN**
MFCC(mean,std,deltamean)	66.60	55.30	MFCC(std,deltamean (2),deltastd(2))	56.60	43.66
MFCC(mean,Std,deltastd)	63.33	57.33	MFCC(deltamean,mean(2),std(2))	67.30	53.00
MFCC(mean,deltamean,deltastd)	67.30	56.67	MFCC(deltamean,mean(2),deltamean(2))	61.00	44.60
MFCC(std,deltamean,deltastd)	68.30	57.00	MFCC(deltamean,mean(2),deltastd(2))	61.60	45.00
MFCC(mean,mean(2),std(2))	66.00	56.33	MFCC(deltamean,std(2),deltamean(2))	68.33	60.33
MFCC(mean,mean(2,deltamean(2))	62.33	53.66	MFCC(deltamean,std(2),deltastd(2))	59.67	52.00
MFCC(mean,mean(2),deltastd(2))	60.00	51.67	MFCC(deltamean,deltamean(2),deltastd(2))	64.00	57.33
MFCC(mean,std(2),deltamean(2))	62.30	56.33	MFCC(deltastd,mean(2),std(2))	63.00	52.33
MFCC(mean,std(2),deltastd2))	62.33	53.66	MFCC(deltastd,mean(2),deltamean(2))	63.67	47.66
MFCC(mean,deltamean(2),deltastd 2))	64.66	57.66	MFCC(deltastd,mean(2),deltastd(2))	64.67	53.00
MFCC(std,mean(2),Std (2))	60.30	48.00	MFCC(deltastd,std(2),deltamean(2))	66.34	50.66
MFCC(std,mean(2),deltamean (2))	54.00	46.67	MFCC(deltastd,std(2),deltastd(2))	61.33	51.33
MFCC(std,mean(2),deltastd (2))	58.30	48.66	MFCC(deltastd,deltamean(2),deltastd(2))	60.00	45.66
MFCC(std,std(2),deltamean (2))	65.90	52.33	MFCC(mean(2),std(2,deltamean(2))	59.70	54.33
MFCC(std,std(2),deltastd (2))	63.00	51.33	MFCC(mean(2),std(2,deltastd(2))	58.19	51.00

Table 3. Four Feature Combination Classification Accuracy (%) for SRC Vs KNN

Intro Piece of Music (All ethnic classes)		
Feature Combination	**SRC**	**KNN**
MFCC(mean,std,deltamean,deltastd)	68.38	59.00
MFCC(mean,std,mean(2),std(2))	65.00	54.66
MFCC(mean,std,mean(2),deltamean(2))	68.34	57.67
MFCC(mean,std,mean(2),deltastd(2))	64.41	53.67
MFCC(mean,std,std(2),deltamean(2))	63.54	58.67
MFCC(mean,std,std(2),deltastd(2))	72.47	62.00
MFCC(mean,std,deltamean(2),deltastd(2))	66.42	60.67
MFCC(deltamean,deltastd,mean(2),std(2))	65.93	49.00
MFCC(deltamean,deltastd,mean(2),deltamean(2))	65.45	47.33
MFCC(deltamean,deltastd,mean(2),deltastd(2))	62.27	50.00
MFCC(deltamean,deltastd,std(2),deltamean(2))	70.69	54.33
MFCC(deltamean,deltastd,std(2),deltastd2))	65.13	55.00
MFCC(deltamean,deltastd,deltamean(2) ,deltastd(2))	70.65	47.67

Table 4. Five Feature Combination Classification Accuracy (%) for SRC Vs KNN

Intro Piece of Music (All ethnic classes)		
Feature Combination	**SRC**	**KNN**
MFCC(mean,std,deltamean, deltastd,mean(2))	66.17	55.33
MFCC(mean,std,deltamean, deltastd,std(2))	70.68	57.00
MFCC(mean,std,deltamean, deltastd,deltamean(2))	66.83	55.67
MFCC(mean,std,deltamean, deltastd,deltastd(2))	63.13	58.33

Table 5. Best Feature Combination Classification Accuracy (%) for intro piece of music

Feature Combination	**SRC**	**KNN**
MFCC(std,deltamean)	75.00	51.33
MFCC(deltamean,std(2))	69.67	59.00
MFCC(deltamean,std(2),deltamean(2))	68.33	60.33
MFCC(mean,deltamean,deltastd)	67.30	56.67
MFCC(mean,std,std(2),deltastd(2))	72.47	62.00
MFCC(deltamean,deltastd,std(2),deltamean(2))	70.69	54.33
MFCC(mean,std,deltamean,deltastd,std(2))	70.68	57.00
MFCC(mean,std,deltamean,deltastd,deltamean(2)	66.83	55.67

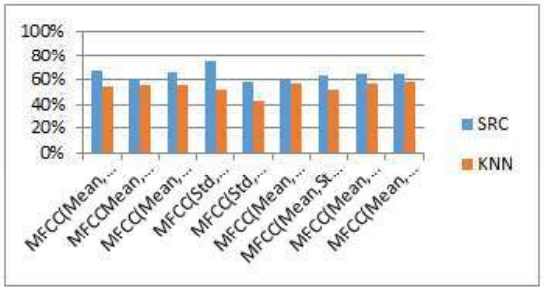

Fig 2. Charts of two features combination for intro piece of music classification

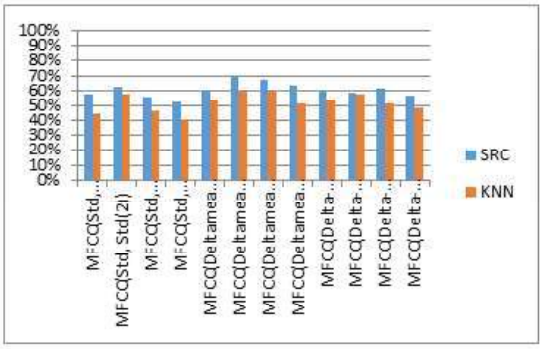

Fig 3. Charts of two features combination for intro piece of music classification

Table 6. Classification accuracy (%) of all features(114) for all ethnic classes

Intro Piece of Music Classification Accuracy (%)		
Ethnic Classes	**SRC**	**KNN**
Kachin	78.00	63.33
Shan	58.00	60.00
Mon	52.00	46.67
Kayin	82.00	65.00
Yakhine	50.00	35.00
All	**64.00**	**54.00**

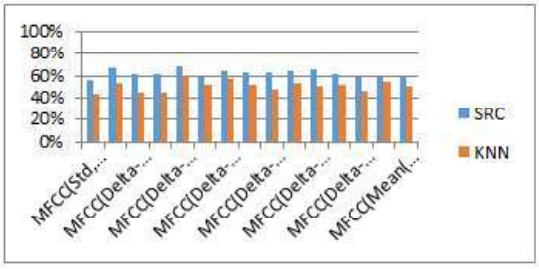

Fig 4. Charts of three features combination for intro piece of music classification

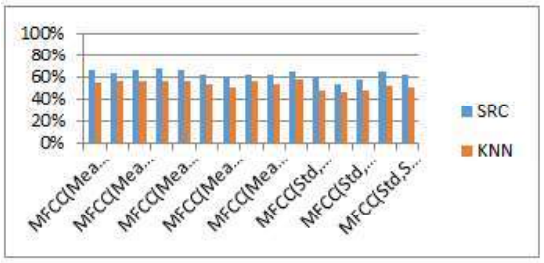

Fig 5. Charts of three features combination for intro piece of music classification

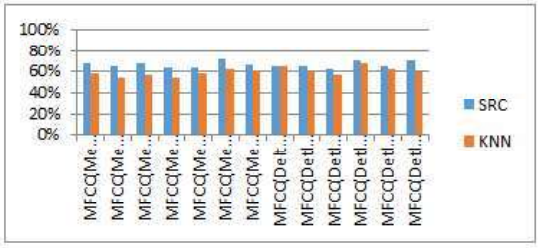

Fig 6. Charts of four features combination for intro piece of music classification

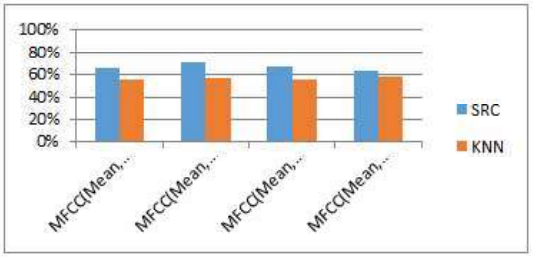

Fig 7. Charts of five features combination for intro piece of music classification

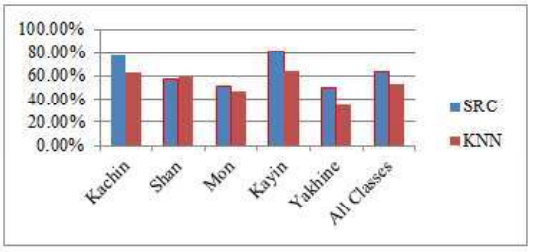

Fig 8. Charts of music classification for all features

ACKNOWLEDMENT

There has been an inspiring, often exciting, sometimes challenging, but definitely always interesting experience from the commencement until the completion of this research. The authors wish to thank my sincere gratitude to my supervisor Dr. Nu War, University of Computer Studies, Mandalay for kind and constant encouragement, close guidance, patient supervision and detail technical a ssistance throughout my research work. Additional thanks are extended to all my friends, for their help and support during my research.

REFERENCES

[1] Y. Liu, Q. Xiang, Y. Wang, and L. Cai, "Cultural style based music classification of audio signals," 2009, pp. 57–60.

[2] S. Jothilakshmi and N. Kathiresan, "Automatic Music Genre Classification for Indian Music," p. 5.

[3] S. H. Deshmukh and D. S. G. Bhirud, "Analysis and application of audio features extraction and classification method to be used for North Indian Classical Music's singer identification problem," vol. 3, no. 2, p. 6, 2014.

[4] G. Madjarov, G. Pesanski, and D. Spasovski, "Automatic Music Classiûcation into Genres," *ICT Innov.*, p. 10, 2012.

[5] M. Banitalebi-Dehkordi and A. Banitalebi-Dehkordi, "Music Genre Classification Using Spectral Analysis and Sparse Representation of the Signals," *ArXiv180304652 Cs Eess*, Mar. 2018.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.