International Journal of Trend in Research and Development, Volume 6(4), ISSN: 2394-9333
www.ijtrd.com

A Classification and Comparison of Feature Extraction For Myanmar Ethnic Songs

May Thu Myint
Doctor of Philosophy in Faculty of Computer Science, Computer University (Hpa-an), Hpa-an, Myanmar

Abstract—At present, digital music collections are commonly accessed through genre, style, mood, release year, and artist. Music classification based on ethnic traditional style is useful for music analysis and has potential applications in retrieval and recommendation systems. The proposed system is designed to classify for Myanmar ethnic music by using sparse representation classification method to define the class label. In this work, the proposed system present the first attempt to classify audio signals automatically according to their ethnic traditional style, which are characterized by timbres features. Ethnic music classification system has three main steps: preprocessing- the audio piece is converted from stereo to mono channel, feature extraction : after preprocessing, the major nine features are extracted such as zcr, centroid, skew, kurtosis, bandwidth, MFCC mean, MFCC std, MFCC delta-mean, MFCC delta- std. After the feature extraction, the final step is classification in which these extracted features are classified by using SRC classifier. For classification purpose, learning supervised classifier will be employed to determine the ethnic traditional style of a music piece. In the proposed system, it is used the whole song datasets that is in evaluation of individual feature experiment and feature combination experiment. Also, the only five Myanmar ethnic groups are considered such as Kachin, Kayin, Mon, Shan and Rakhine ethnic songs. In whole song experiment, it used 90 songs for training and testing processes and then classification of overall average accuracy of the five ethnic classes from the 114 features receives 65%. Moreover, in whole song experiment, the accuracies of Kachin, Shan , Mon, Kayin, Yakhine ethnic song receive 50%, 81.64%, 51.68%, 71.66%, 70% respectively.

Keywords—*Myanmar Ethnic Music; Sparse Representation Classifier; Preprocessing;,Feature Extraction;Spectral Shape Feature*

I. INTRODUCTION

Culture has a significant influence on music in term of creation, performance and interpretation. People from a certain cultural background often prefer music with a particular cultural style. Therefore cultural style information is useful for music browsing, retrieval and recommendation. This is important in today’s Internet enabled global village where people can perform cross-culture music exploration. The rapid growth in computational power available in personal computers, the development of high-speed digital communication networks, high-density storage devices and the availability of input/output devices for data from different media has led to a sudden explosion in the amount of data handled by computers. The above advances in technology have led to an increased use of non-textual data such as video, audio, images and graphics in many applications. These non-traditional data types which are related and handled together is popularly known as multimedia data and can provide more effective dissemination of information and communication of ideas in science, engineering, medicine, biology, social science and commerce. It can be observed that music pieces with the same cultural style share similar attributes such as the tuning system, musical scale and instrumentation. Based on this observation, music audio signals can be classified according to their cultural styles using machine learning methods.

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. Automatic classification of musical style is gaining more and more importance since it may serve as a way to structure and organize the increasingly large number of music files available on the Web. For data to be gainfully and meaningfully used in various applications, it is essential to have efficient schemes for data management and manipulation which broadly involve acquisition, organization, storage, query, retrieval, transmission, and presentation of data. Automatic music classification can be applied to a wide variety of tasks, both academic and commercial in nature. However, audio applications would also increase in number as multimedia databases proliferate and as specialized audio applications gain currency. Audio could serve as and independent data type or as part of multimedia data. These specialized applications originate in domains such as the medical industry (ultrasound, electrocardiography), entertainment industry (music, speed recordings and sound databases for virtual reality environments) and security (voice recognition, sound identification). A major challenge in this field is the automatic classification of audio and music [91, 79, 81, 80, 34, 92, 22]. Most audio classification systems combine two processing stages: feature extraction followed by classification. A variety of signal features have been used for this purpose, including low-level parameters such as the zero-crossing rate, signal bandwidth, spectral centroid, and signal energy. There has also been some recent work on automatic music genre detection. Several different classification strategies have been employed in these studies, including multivariate Gaussian models, Gaussian mixture models, self-organizing maps, neural networks, k-nearest neighbor schemes and hidden Markov models. In some cases, the classification scheme does not influence the classification accuracy [76, 83] suggesting that the topology of the feature space is relatively simple. An important implication of these findings is that, perhaps further advances could be made by developing more powerful features or at least understanding the feature space, rather than building new classification schemes.

II. PRELIMINARIES

A. *Audio Dataset*

The audio files 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 90 songs from the popular ethnic music songs categorized as 18 audio recordings of each ethnic classes respectively. These songs are from MRTV radio station. Each music piece lasts about 3 to 5 minutes in length. From all of the five ethnic classes, experiment I, the whole duration of music used for individual feature comparison and in experiment II consists of the whole duration of music is used for evaluating the feature combination in SRC classifier and the best feature combination are expressed in comparison with SRC classifier.

B. System Architecture

All audio based Myanmar ethnic music classification system can be divided into three main processes. The first process is that of the input audio is firstly preprocessed and musical feature extraction on the different musical songs. The resulting features are used whenever it wants to expand the feature set of known songs or to query the feature set when the song is unknown. The final process is that of finding the classify the input musical signal songs based on the training feature set as well as testing,the of the system to define the class label from input audio music query. In the preprocessing stage, the input audio piece is converted from stereo to mono channel. The audio sample (sampling of 44 kHz) is framing with 100 ms. These audio sample frames was taken 50% overlapping frame between the successive frames. Preprocessing method has some consequence effects to the feature extraction and classification of the system. So, it is important to use a very suitable method. All of 114 features are extracted after preprocessing stag which all of features and each feature are used by Sparse Representation Classifier (SRC) for classifying each ethnic class. According to the nature of SRC is that the test data is considered as a liner composition of the training data set belonging to the same category if sufficient training samples are available for each class. SR coefficient vector will have only a few significant coefficients.

III. RESEARCH METHODS

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, theses audio samples are extracted the timbre features (zcr, centroid, shew, centroid kurtosis bandwidth, MFCC mean, MFCC std, MFCC delta-mean, MFCC delta std) which may be calculated based on the basic samples after preprocessing., According to the point of view, the extraction processes of the feature are: Features that are directly computed on the audio samples data as, for example, zero-crossing rate (the rate that the waveform changes from positive to negative values), spectral centroid (the "gravity centre" of the spectrum, Skewness- measures the degree of a symmetry exhibited by the data, Spectral centroid- mid-point of the spectral power distribution of a signal, Kurtosis- characterizes the relative peakedness of a distribution compared to normal distribution, Bandwidth- Rate of data transfer, bit rate or throughput, measured in bits per second (bit/s), Mel frequency cepstral coefficients (MFCC)- natural to use the mel-scale and log amplitude ,since it relates to how perceive sounds. Finally, after feature extraction, these extracted features are classified by using the Sparse Representation Classifier (SRC) to define each ethnic class label. In the chapter, it has been presented the timbre features and SRC classifier based on ethnic songs based on Time-domain and Frequency-domain analysis.

A. Feature Description

Summarizing here the main features used in musical classification as following:

- **Temporal features:** features computed from the audio signal frame (zero-crossing rate, linear prediction coefficients, etc.).
- **Energy features:** features referring to the energy content of the signal (Root Mean Square energy of the signal frame, energy of the harmonic component of the power spectrum, energy of the noisy part of the power spectrum, etc.).
- **Spectral shape features:** features describing the shape of the power spectrum of a signal frame: centroid, spread, skewness, kurtosis, slope, roll-off frequency, variation, Mel-Frequency Cepstral Coefficients (MFCCs).
- **Perceptual features**: features computed using a model of the human earring process (relative specific loudness, sharpness, spread). These features can be categorized into short time features and long time features.

The short time features are computed at regular time intervals, over short windows of typical length between 10 and 60ms. These are mainly based on spectrum-derived quantity within a short segment. In contrast, the long-time features are computed over larger windows commonly called as texture windows. The typical length of texture window is 1 second. These features characterize the variation of spectral shape or beat information over a long segment [50, 95].

B. Feature Extraction

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 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. Timbre is the property which is second-most developed because an understanding of timbre is needed for synthesizers, and because timbre features have proved useful in other fields. The major aims of the timbre perception studies are to develop a theory of sounds classification. In the study, the author focused on experimenting different sounds coming from various orchestra instruments. According to this, timbre content is the common music feature that is used to distinguish different aspects of the music and instrumentations, and if they are combined with other features such as rhythm and tonal characteristics, they usually are enough to discriminate various styles of the music. The core timbre features used in Myanmar ethnic music classification system are described in the following subsection.

C. Feature Used

1) Mel Frequency Cepstral Coefficients: The Mel frequency cepstral coefficients (MFCCs) are derived from the Fourier transform of the audio signal adapted to the Mel scale, which is a perceptually- determined pitch scale that corresponds to human hearing. Below about 1000 Hz, humans perceive a doubling of frequency as a doubling of pitch, but above this point the association becomes logarithmic. The Mel scale approximates thisrelationship as shown in the following conversion between frequencies in Hz (f) and Mels (m**);**

$$MEL(F_M) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

$$MFCC_I = \sum_{k=0}^{N-1} X_K \cos\left[I\left(K+\frac{1}{2}\right)\frac{\pi}{2}\right];\ I=0,1,\ldots,M-1 \quad (2)$$

where M is the number of desired cepstral coefficients, N is the number of filters, and X_k is the log power output of the k th filter. Each frame of signals in time domain will be represented by a vector of 13 features. The MFCCs are determined by applying a Discrete Consine Transform (DCT) to each of a series of Mel filter bank energies, resulting in a series of coefficients, of which the first thirteen are used. It is reasonable to suppose that subtle differences in instruments might be detected with the use of more coefficients, so thirteen were extracted for these experiments. Mel Frequency Cepstral Coefficients (MFCC) are a compact representation of the spectrum of an audio signal that takes into account the nonlinear human perception of pitch.

2) Centroid: The Spectral Centroid is the center of gravity of second moment of the spectrum. It provides a measure of the shape of a spectrum; higher values have more high- frequency components and vice-versa. Many types of music involve percussive sounds which push the spectral mean higher by including high-frequency noise. The spectral centroid for a frame is computed as follows: (3)

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

where the ratio of the sum of spectral magnitude M[f]weighted by frequency f to the sum of spectral magnitude.

3) Time Domain Zero- Crossings: Zero- crossing is simply a count of the number times the signal crosses from positive to negative of vice-vesa. It is a time-domain feature, and as such very fast to calculate. A zero-crossing occurs when successive samples in a digital signal have different signs. Therefore, the rate of zero-crossings can be used as a simple measure of a signal's frequency content. For simple signals, the ZCR is directly related to the fundamental frequency (f_0). First of all, it is very hard to determine the fundamental frequency of a complex auditory scene containing several sound sources (instruments, voices, and sound effects). 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. ZCR curves to distinguish environmental sounds based on regularity, periodicity, and range of amplitude. The ZCR curves are calculated as follows:

$$Z_n = \sum_m \left| sgn[x(m)] - sgn[x(m-1)]w(n-m) \right| \quad (4)$$

$$sgn\,[x(n)] = \begin{cases} -1, & x(n) \geq 0 \\ 1, & x(n) \leq 0 \end{cases} \quad (5)$$

$$w(n) = \begin{cases} 1/2 & 0 \leq n \leq 0 \\ 0 & otherwise \end{cases} \quad (6)$$

where m is the window size in this short-time function.

4) Skewness: Importantly, the skewness does not determine the relationship of mean and median.

a) Negative skew: The left tail islonger; 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.

b) 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.

$$skewness = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{ns^3} \quad (7)$$

where $\bar{x}$ the sample mean and s is is the sample standard deviation and n is the number of samples.

5) Kurtosis: It is basically because kurtosis represents a movement of mass that does not affect the variance. Consider the case of positive kurtosis, where heavier tails are often accompanied by a higher peak. In Fig. 1, note that if mass is simply moved from the shoulders of a distribution to its tails, then the variance will also be larger.

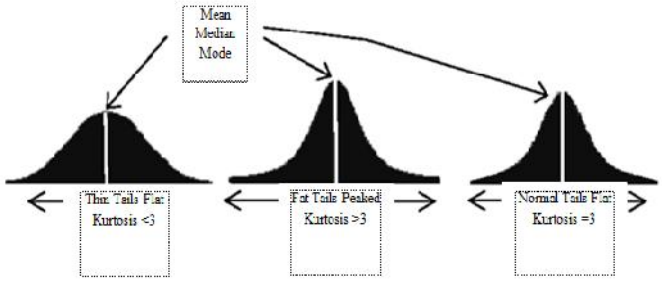

Fig. 1. Illustrative prototype histograms

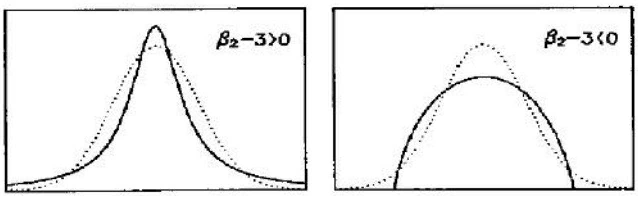

Fig. 2. Illustrative of kurtosis

For negative kurtosis, the variance will be unchanged if mass is moved from the tails and center of the distribution to its shoulders, thus resulting in light tails and flatness. In Fig.2, the dotted lines show normal distributions, whereas the solid lines show distributions with positive kurtosis (left panel) and negative kurtosis (right panel).

6) Bandwidth: Bandwidth (signal processing) or analog bandwidth, frequency bandwidth or radio bandwidth: a measure of the width of a range of frequencies, measured in hertz. Alternatively, bandwidth is the difference between the upper and lower frequencies in a continuous set of frequencies. It is typically measured in hertz, A key characteristic of bandwidth is that any band of a given width can carry the same amount of information, regardless of where that band is located in the frequency spectrum.

D. Sparse Representation Classifier

Sparse representations have become a widely popular approach in the context of remote sensing data processing [12]. Such representations are widely used in many areas [61], including spectral unmixing and classification. These approaches intend to represent most observations or image pixels with linear combinations of a small number of elementary (dictionary) samples, which are often called *atoms*, chosen from an over complete training dictionary. Formally, an over complete dictionary is a collection of atoms, such that the number of atoms exceeds the dimension of the image space, and any image pixel can be represented by more than one combination of different atoms. In order to identify an audio signal, its features need to be compared to the known features contained in a feature set. The SRC classifier algorithm is that the test data is considered as a linear composition of the training data set belonging to the same category if sufficient training samples are available for each class. For the very large dataset, the SRC classifier is appropriate and thus optimization of the system is crucial. SRC first encodes a query sample as a linear combination of a few atoms from a pre-defined dictionary. It then identifies the label by evaluating which class results in the minimum reconstruction error. The effectiveness of SRC is limited by an important assumption that data points from different classes are not distributed along the same radius direction. SRC can accurately classify test samples outside the overlapped region but within that region, the accuracy is close to random guess.

TABLE I. SPARSE REPRESENTATION CLASSIFIER

Input: matrix of training samples $A=[A_1, A_2, \ldots, A_k] \in R^{m\times n}$ for k classes; testing sample $y \in R^m$

Step1: Normalize the columns of A to have unit L_0- norm.

Step 2: Extract the meta- samples of every class using NMF.

Step 3: Solve the optimization problem defined in

$$J(x,\lambda) = \min_x\{\ \| W_x - y \| + \lambda \| x \|_1 \}$$

Step 4: Compute the residuals $r_i(y) = \| y - W\delta_i(x) \|_2$

Output: Identify $= \underset{i}{\mathrm{argmin}}\ r_i(y)$

Fig. 3. Illustrative of sparse representation classifier

Consider a training gene expression dataset represented by an $m\times n$ matrix A with m genes and n samples. Since the microarray data typically contain thousands of genes on each chip, and the number of collected tumor samples is much smaller than that of genes, we have m$\gg$ n. Arranging the n^i samples of the i-th class ($1\leq i\leq k$) as a matrix $A_i = [c_{i,1}, c_{i,2}, \ldots, c_{i,ni}] \in IR^{m\times n_i}$, with each sample being a column. Given that the training samples of the i[th] class are sufficient, any new (testing) sample $y \in IR^m$ in the same class will approximately lie in the linear span of the training samples associated with class i [18]:

$$y = \alpha_{i,1} c_{i,1} + \alpha_{i,2} c_{i,2} + \ldots \alpha_{i,n\,i} c_{i,ni} \qquad (8)$$

for some scalars $\alpha_{i,j} \in IR\, j= 1,2,\ldots,n_i$.

The membership i of the new testing sample y is unknown. We arrange the training data samples of each class in matrix A . Suppose that the samples with the same class are conjoint, i.e., A= $[A_1, A_2, \ldots, A_k]$, then the linear representation of y can be rewritten in terms of all the training samples as y= Ax_0 where, ideally,

$$x_0=[0,\ldots,0,\ \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,ni}, 0,\ldots,0]^T \in IR^n \qquad (9)$$

is a coefficient vector whose entries are zero except for those associated with the i[th] class. In other words, the nonzero entries in the extrimate x_0 will be associated with the columns of A from a single object class i so that we can assign the testing sample y to that class. From Eq. (8) can see that the representation of y is naturally sparse if the number of object classes k is large. The problem can be converted into how to find a column vector x such that y = Ax_0 and x is minimized, where $\|x\|_0$ is the l_0-norm of x and it is equivalent to the number of nonzero elements in vector x, i.e., the so-called sparse representation (SR). It can be expressed as the following optimization problem:

$$x_0=\mathrm{argmin}\|x\|_0 \ \text{ subject to } A_x=y \qquad (10)$$

The solution to the L_0 – minimization problem in Eq. (10) is equivalent to the solution to the following 1_1–minimization problem:

$$x_1 = \mathrm{argmin}\ \|x\|_0 \ \text{ subject to } A_x = y \qquad (11)$$

This problem can be solved in polynomial time by standard linear programming methods [12]. For A, whose size m >> n, Eq. (11) does not have exact solutions. To solve this problem, generalized versions of Eq. (11) is considered which allows for certain degree f noise, i.e., find a vector x such that the following objective function is minimized:

$$J(x,\lambda) = \min_x\{\ \|Ax-y\|_2 + \lambda\ \|x\|_1\} \qquad (12)$$

Using this, Eq. (4.20) is reduced to solving an l_1 -regularized least square problem.The positive parameter λ in Eq. (13) is a scalar regularization that balances the reconstruction error and sparsity. This optimization problem can also be solved by standard linear programming be methods [12].

$$\min_i r_i(y)=\min_i\{\ \|\ y - W\delta i\ (x)\ \|_2\ \} \qquad (13)$$

For each class i, let δ_i, $IR^n \longrightarrow IR^n$ be the characteristic function which selects the coefficients associated with the i[th] class. For $x \in IR^n$, is a vector whose nonzero entries are the ones from class i in x. Using only the coefficients from the i[th] class, one can reconstruct the given test sample y as $y_i = W\delta_{i(x)}$. We then classify based on these approximations by assigning it to the class that minimizes the residual between y and $\hat{y}$.

IV. EXPERIMENTAL SETUP

The performance of the proposed Myanmar ethnic music classification system evaluates the features in the various features methods and classifying the similar structure of a set of music pieces. The Sparse Representation Classifier is performed against the ethnic music system based on the extracted timbral features. The SRC is that the test data is considered as a linear composition of the training data set belonging to the same category if sufficient training samples are available for each class. All of terms defined in below that are used in all experiment result's tables and graphs. Zcr, Centroid, Bandwidth, Skew, Kurtosis, MFCC mean, MFCC std, MFCC delta-mean and MFCC delta-std features are expressed in the results of calculating mean value of these features and Zcr(2), Centroid(2), Bandwidth(2), Skew(2), Kurtosis(2), MFCC mean(2), MFCC std(2), MFCC delta-mean(2) and MFCC delta-std (2) features are expressed in that results of calculating standard deviation value of theses feature.

TABLE II. LIST OF FEATURES NAME

Zcr	=	mean(Zcr)
Centroid	=	Mean(Centroid)
Bandwidth	=	mean(Bandwidth)
Skew	=	mean(Skew)
Kurtosis	=	mean(Kurtosis)
MFCC mean	=	mean(MFCC mean)
MFCC std	=	mean(MFCC std)
MFCC Delta-mean	=	mean(MFCC Delta-mean)
MFCC Delta-std	=	mean(MFCC Delta-Std)
Zcr (2)	=	std(Zcr)

Centroid (2)	=	std(Centroid)
Bandwidth (2)	=	std(Bandwidth)
Skew (2)	=	std(Skew)
Kurtosis (2)	=	std (Kurtosis)
MFCC mean (2)	=	std (MFCC mean)
MFCC std (2)	=	std (MFCC std)
MFCC Delta- mean(2)	=	std (MFCC Delta-mean)
MFCC Delta-Std (2)	=	std (MFCC Delta-Std)

A. Evaluation of Individual Feature

In this experiment, the proposed system is evaluated with the individual feature type. There are major nine features in this system. So, the zcr(mean), zcr(std), centroid(mean), centroid(std), bandwidth(mean), bandwidth(std), MFCC mean(mean), MFCC std(mean), MFCC std(mean), MFCC std(std), MFCC delta-mean(mean), MFCC delta-mean(std), MFCC delta-std(mean) and delta-std(std) are tested in this part. All of ethnic classes are Kachin, Shan, Mon, Kayin and Yakhine ethnic songs.The evaluation results are shown in following subsection.

In table III, the mean value of bandwidth is the highest accuracy of 90.00% are achieved for Kachin ethnic songs and the results of Zcr is the good accuracy than other features for Shan ethnic class. In the individual feature experiment, the accuracy of MFCC mean, MFCC delta-mean and MFCC std give the best accuracy of between 70% and 80% for Mon, Kayin and Yakhine ethnic songs. For all ethnic classes, MFCC delta-mean is the best feature and its classification accuracy is 72.00% among the whole duration songs. The centroid, skew, kurtosis feature give the low accuracy than other features among these five ethnic classes.

In this table IV, the result value of (MFCC delta-std) is the highest accuracy of 100.00% are achieved for Kachin ethnic songs and the results of std (MFCC delta-mean) is the good accuracy than other features for Shan ethnic class and its accuracy is 76.66%.. In the individual feature experiment, the MFCC std(2) feature give the best accuracy of between 80% and 90% for Kayin and Yakhine ethnic songs and this feature give the low accuracy of 56.66% for Mon ethnic songs. Also, for all ethnic classes, MFCC Std(2) is the best feature that the classification accuracy is 76.33% and the accuracy result is higher than the mean value of MFCC delta-mean expressed in table II.

TABLE III. CLASSIFICATION ACCURACY (%) FOR MEAN VALUE OF INDIVIDUAL FEATURE

Classification Accuracy (%) for mean value of individual feature									
Ethnic Class	**Zcr**	Centroid	Bandwidth	Skew	Kurtosis	MFCC mean	MFCC std	MFCC delta-mean	MFCC delta-std
Kachin	**20.00%**	**35.00%**	**90.00%**	**0.00%**	**75.00%**	**75.00%**	**78.33%**	**71.66%**	**76.66%**
Shan	**76.67%**	**10.00%**	**0.00%**	**31.66%**	**10.00%**	**48.33%**	**56.66%**	**60.00%**	**60.00%**
Mon	**0.00%**	**0.00%**	**8.33%**	**10.00%**	**18.33%**	**78.33%**	**47.00%**	**65.00%**	**61.66%**
Kayin	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**0.00%**	**75.00%**	**71.66%**	**86.66%**	**61.66%**
Yakhine	**10.00%**	**48.33%**	**19.66%**	**56.67%**	**0.00%**	**65.00%**	**80.00%**	**76.66%**	**55.00%**
All Ethnic Classes	**21.33%**	**18.67%**	**19.66%**	**18.99%**	**20.66%**	**68.33%**	**67.99%**	**72.00%**	**62.99%**

TABLE IV. CLASSIFICATION ACCURACY (%) FOR STD VALUE OF INDIVIDUAL FEATURE

Classification Accuracy (%) for std value of individual feature									
Ethnic Class	**Zcr (2)**	Centroid(2)	Bandwidth (2)	Skew(2)	Kurtosis(2)	MFCC mean (2)	MFCC Std (2)	MFCC delta-mean (2)	MFCC delta-std (2)
Kachin	**58.33%**	**28.33%**	**91.66%**	**61.66%**	**76.66%**	**81.66%**	**94.99%**	**91.66%**	**100.00%**
Shan	**0.00%**	**53.33%**	**6.66%**	**18.33%**	**0.00%**	**63.33%**	**63.33%**	**76.66%**	**68.33%**
Mon	**26.66%**	**21.66%**	**0.00%**	**0.00%**	**30.00%**	**55.00%**	**56.66%**	**36.66%**	**15.00%**
Kayin	**0.00%**	**10.00%**	**0.00%**	**0.00%**	**0.00%**	**68.33%**	**86.66%**	**63.33%**	**25.00%**
Yakhine	**10.00%**	**10.00%**	**0.00%**	**25.00%**	**0.00%**	**59.99%**	**80.00%**	**76.66%**	**10.00%**
All Ethnic Classes	**19.00%**	**20.33%**	**19.66%**	**20.99%**	**21.33%**	**65.67%**	**76.33%**	**69.00%**	**43.66%**

B. *Evaluation of feature combination results*

The following classification results are described the feature combination results by combining each feature for all ethnic classes . In comparison, these features combination are used from Sparse Representation Classifier to classify all ethnic song classes for whole duration of music experiment.

According to the table (V), feature combination of MFCC std, and MFCC delta-mean are tested on all of five ethnic classes which is the best classification results of 73.33% accuracy from SRC classifier for all ethnic songs. Also, the feature combination of MFCC mean, and MFCC delta-mean (2) are the second best feature combination on all five ethnic classes in which these classification accuracy is achieved 72% accuracy. The feature combination of MFCC (Std, Std(2)) discriminate the highest accuracy of 75.33% than other feature combination from SRC classifier. Also, other feature combination are achieved the classification rate of between 60% and 80% for all ethnic classes.

In table (VI), using the three feature combination of MFCC (Delta-mean, Std(2), Delta-mean(2)), it discrimate 79.01% of highest accuracy for all ethnic songs than the best accuracy of two feature combination which is expressed in table V and MFCC(Mean(2), Std(2, Delta-mean(2)) achieves 78.61% of high accuracy for all ethnic songs. According the this table, this MFCC (Std, Std(2),Delta-mean (2)) feature combination give the lower accuracy result of 63.00% than other three feature combination.

. TABLE V. TWO FEATURE COMBINATION CLASSIFICATION ACCURACY (%)

Whole duration of music (All ethnic classes)			
Classification Accuracy (%)			
Feature Combination	**SRC**	**Feature Combination**	**SRC**
MFCC mean, MFCC Std	69.00%	MFCC(Std, Std(2))	75.33%
MFCC mean, MFCC delta-mean	65.00%	MFCC(Std, Mean(2))	70.00%
MFCC mean, MFCC delta-std	70.67%	MFCC(Std, Delta-mean(2))	70.33%
MFCC(Std, delta-mean)	73.33%	MFCC(Std, Delta-std(2))	63.33%
MFCC(Std, delta-std)	66.67%	MFCC(Delta-mean, Mean(2))	71.33%
MFCC(Mean, Mean(2))	68.67%	MFCC(Delta-mean, Std(2))	70.33%
MFCC(Mean, Std(2))	65.00%	MFCC(Delta-mean, Delta-mean(2))	69.67%
MFCC(Mean, Delta-mean(2))	72.00%	MFCC(Delta-mean, Delta-std(2))	69.00%
MFCC(Mean, Delta-std(2))	70.33%		

. TABLE VI. THREE FEATURE COMBINATION CLASSIFICATION ACCURACY (%)

Whole Duration of Music (All ethnic classes)			
Classification Accuracy (%)			
Feature Combination	**SRC**	**Feature Combination**	**SRC**
MFCC(Mean, Std, Delta-mean)	68.67%	MFCC(Delta-mean, Std(2), Delta-mean(2))	79.01%
MFCC(Mean, Std, Delta-std)	66.33%	MFCC(Delta-mean, Std(2), Delta-std(2))	78.00%
MFCC(Mean, Delta-mean, Delta-std)	64.47%	MFCC(Delta-mean, Delta-mean(2), Delta-std(2),)	77.00%
MFCC(Std, Delta-mean,Delta-std)	74.67%	MFCC(Delta-std, mean(2), std(2))	75.67%
MFCC(Mean, Mean(2), Std(2))	73.33%	MFCC(Delta-std, mean(2), Delta-mean(2))	67.00%
MFCC(Mean, Mean(2,Delta-mean(2))	71.67%	MFCC(Delta-std, mean(2), Delta-std(2))	69.66%
MFCC(Mean, Mean(2),Delta-std(2))	66.00%	MFCC(Delta-std, Std(2), Delta-mean(2))	76.67%
MFCC(Mean, Std(2),Delta-mean(2))	71.00%	MFCC(Delta-std, Std(2), Delta-std(2))	72.33%
MFCC(Mean, Std(2),Delta-std2))	67.00%	MFCC(Delta-std, Delta-mean(2), Delta-std(2))	77.00%
MFCC(Mean, Delta-mean(2),Delta-std2))	70.44%	MFCC(Mean(2), Std(2, Delta-mean(2))	78.61%
MFCC(Std, Mean(2),Std (2))	74.33%	MFCC(Mean(2), Std(2, Delta-std(2))	73.84%
MFCC(Std, Mean(2),Delta-mean (2))	71.66%	MFCC(Std, Delta-mean (2),Delta-std(2))	67.33%

MFCC(Std, Mean(2),Delta-std (2))	73.00%	MFCC(Delta-mean, mean(2), std(2))	73.33%
MFCC(Std, Std(2),Delta-mean (2))	63.00%	MFCC(Delta-mean, mean(2), Delta-mean(2))	73.00%
MFCC(Std, Std(2),Delta-std (2))	76.67%	MFCC(Delta-mean, mean(2), Delta-std(2))	71.67%

. TABLE VII. FOUR FEATURE COMBINATION CLASSIFICATION ACCURACY (%)

Whole Duration of Music (All ethnic classes)			
Classification Accuracy (%)			
Feature Combination	SRC	Feature Combination	SRC
MFCC(Mean. Std, Delta-mean, Delta-std)	61.38%	MFCC(Delta-mean, Delta-std, Mean(2), Std(2))	72.79%
MFCC(Mean, Std, Mean(2), Std(2))	70.13%	MFCC(Delta-mean, Delta-std, Mean(2), Delta-mean(2))	68.45%
MFCC(Mean, Std, Mean(2), Delta-mean(2))	70.95%	MFCC(Delta-mean, Delta-std, Mean(2), Delta-std(2))	69.85%
MFCC(Mean. Std. Mean(2), Delta-std(2))	71.58%	MFCC(Delta-mean, Delta-std, Std(2), Delta-mean(2))	76.17%
MFCC(Mean, Std, Std(2), Delta-mean(2))	71.94%	MFCC(Delta-mean, Delta-std, Std(2), Delta-std2))	74.40%
MFCC(Mean, Std, Std(2), Delta-std(2))	75.54%	MFCC(Delta-mean, Delta-std, Delta-mean(2), Delta-std(2))	64.81%
MFCC(Mean, Std, Delta-mean(2), Delta-std(2))	72.08%		

. TABLE VIII. FIVE FEATURE COMBINATION CLASSIFICATION ACCURACY (%)

Whole Duration of Music (All ethnic classes)	
Classification Accuracy (%)	
Feature Combination	SRC
MFCC(Mean, Std, Delta-mean, Delta-std, Mean(2))	77.08%
MFCC(Mean, Std, Delta-mean, Delta-std, Std(2))	67.95%
MFCC(Mean, Std, Delta-mean, Delta-std, Delta-mean(2))	75.44%
MFCC(Mean, Std, Delta-mean, Delta-std, Delta-std(2))	72.87%

In table(VII), feature combination of MFCC(Delta-mean, Delta-std, Std(2), Delta-mean(2))are tested on all of five ethnic classes which is the best classification results of 76.17% accuracy from SRC classifier but these result is slightly decrease than the classification results of 79.01% from two feature combination in table (V). However, in this three feature combination table, MFCC (Mean. Std, Delta-mean, Delta-std) feature combination are achieved the low accuracy by using SRC classifier in compared with other feature combination.

According to the table (VIII), feature combination of MFCC (mean, std, delta-mean, delta-std, mean (2)) are tested on all of five ethnic classes which is the best classification results of 77.08% accuracy from SRC classifier. Also, the feature combination of MFCC(Mean, Std, Delta-mean, Delta-std, Delta-mean(2)) and MFCC(Mean, Std, Delta-mean, Delta-std, Delta-std(2)) get the classification accuracy of 75.44% and 72.87% but these MFCC(Mean, Std, Delta-mean, Delta-std, Std(2)) feature combination is slightly lower among the five feature combination experiment.

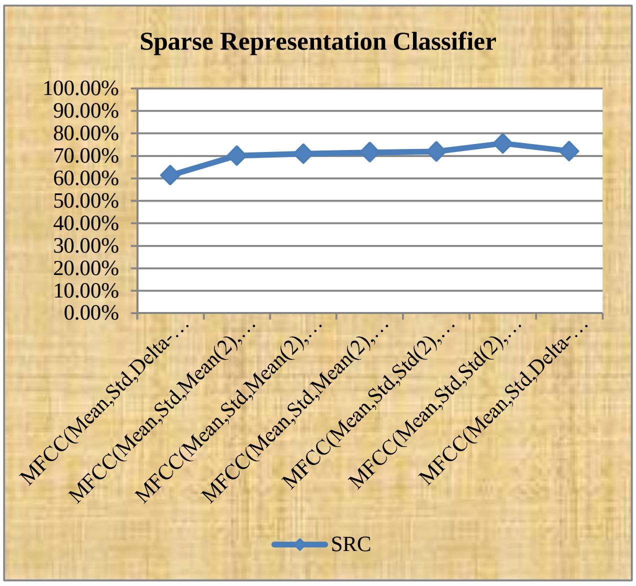

Fig. 4. Charts of four feature combination with SRC

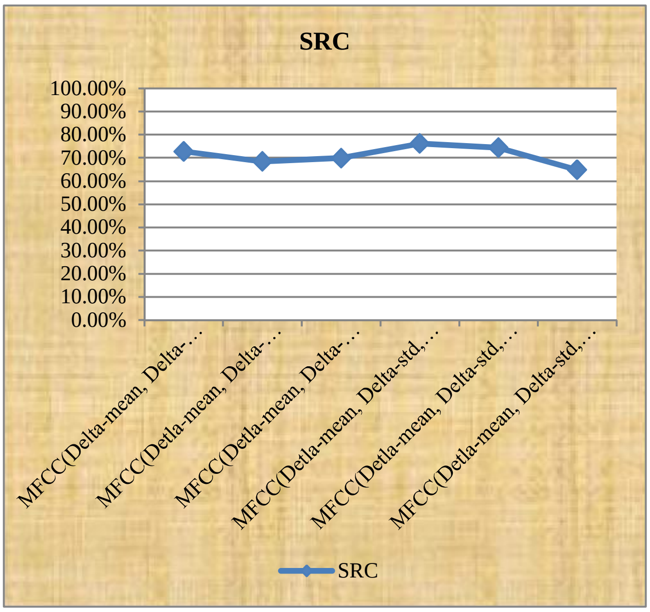

Fig.4. Charts of four feature combination with SRC

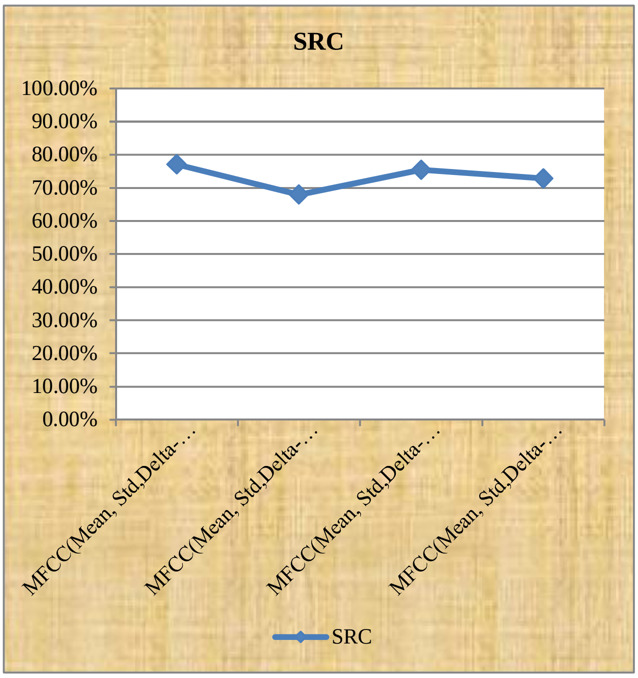

Fig.5. Charts of five feature combination with SRC

Finally, the best feature combinations are selected from all of feature combinations which are used in SRC. According to the table (IX), two feature combination of MFCC (std, std(2)) and MFCC (std,delta-mean) are tested on all of five ethnic classes in which the classification results of 75.33% and 73.33% accuracies from SRC classifier for all ethnic songs. Also, the three feature combination of MFCC (delta-mean, std(2), delta-mean(2)) and MFCC (delta-mean, std(2), delta-std(2)) are used on all five ethnic classes that is these feature combination are the higher classification accuracies of 79.01% and 78% respectively.

. TABLE IX. BEST FEATURE COMBINATION CLASSIFICATION ACCURACY (%)

Classification Accuracy (%)	
Feature Combination	**SRC**
MFCC(Std, Std(2))	75.33%
MFCC(Std, Delta-mean)	73.33%
MFCC(Delta-mean, Std(2), Delta-mean(2))	79.01%
MFCC(Delta-mean, Std(2), Delta-std(2))	78.00%
MFCC(Mean, Std, Std(2), Delta-std(2))	75.54%
MFCC(Delta-mean, Delta-std, Std(2), Delta-mean(2))	76.17%
MFCC(Mean, Std, Delta-mean, Delta-std, Mean(2))	77.08%
MFCC(Mean, Std, Delta-mean, Delta-std, Delta-mean(2))	75.44%

TABLE X. CLASSIFICATION ACCURACY (%) OF WHOLE DURATION OF MUSIC FOR ALL FEATURES

Whole Duration of Music Classification Accuracy (%)	
Ethnic Class	**SRC**
Kachin	50.00%
Shan	81.64%
Mon	51.68%
Kayin	71.66%
Yakhine	70%
All ethnic classes	65%

Table X show the results of whole duration of music by using all features (114) with SRC. The overall results for all ethnic classes are 65% is achieved from SRC at highest overall accuracy for all features. In this table, all features give the best accuracy for Shan ethnic songs among other ethnic classes.

CONCLUSION

Myanmar Ethnic Music classification system is classify the Kachin, Kayin, Shan, Mon and Yakhine ethnic group songs for all experiments using sparse representation classifier (SRC). The system analyses the characteristics of ethnic music and is able to extract timbre features from audio samples that allow it to classify among variety of ethnic songs to label the respective ethnic class. Then, the system labels the input audio data into the corrected ethnic songs. According to the individual feature experiment, the system works with a high degree of accuracy of SRC, Kachin ethnic class is the best for whole duration of music experiment with the MFCC delta-std (2). Therefore, in whole duration of music, MFCC (delta-mean, std(2), delta-mean(2)) is the best three feature combination from SRC. In conclusion, all features testing give the best accuracy for Shan ethnic songs among other ethnic classes. In the classification of ethnic songs, if these songs haven't the real cultural style, the classification can't get the better accuracy. The system can only extract the major nine features (zcr, centroid, bandwidth, skewness, kurtosis, MFCC mean, MFCC std, MFCC delta-mean, MFCC delta-std).

Acknowledgment

There has been an inspiring, often exciting, sometimes challenging, but definitely always interesting experience from the commencement until the completion of this research. It has been made possible by some people, who have supported me. I would like to express my sincere gratitude to my supervisor Dr. Nu War, Head of Hardware Department, Computer University (Mandalay) for kind and constant encouragement, close guidance, patient supervision and detail technical assistance throughout my research work. Additional thanks are extended to all my friends, for their help and support during my studies.

References

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] B. Song, J.Li, M. D. Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, " Remotely Sensed Image Classification Using Sparse Representations of Morphological Attribure Profiles", IEEE, Transactions on geoscience and remote sensing ,Vol.52, No.8, Auguest, 2014.

[3] S.Jothilakshmi and N.Kathiresan," Automatic Music Genre Classification for Indian Music , Department of information technology, International Conference on Software and Computer Applications (ICSCA, 2012)

[4] R. B.Dannenberg, and N. Hu. 2002. Pattern discovery techniques for music audio. Proceedings of the International Symposium on Music Information Retrieval. 63–70.

[5] W. Chai and B. Vercoe, " Folk Music Classification Using Hidden Markov Models", USA,2009

[6] Z. W. Bond, " Unsupervised Classification of Music Signals: Strategies Using Timbre and Rhythm", November , Virginia, 2006.

[7] P.i Toiviainen and T. Eerola, "A method for comparative analysis of folk musci based on musical feature extraction and neural networks", International Symposium on Systematic and Comparative Musicology III International Conference on Cognitive Musicology",2001.

[8] Y. Liu, Q. Xiang, Y. Wang, L. Cai, " Cultural style based music classificaiton of audio signals", Singapore, 2009.

[9] A.Oppenheim, and R.Schafer, Discrete-Time Signal Processing. Prentice Hall.Edgewood Cliffs, NJ. 1989.

[10] B.T.Logan and S.Chu, " Music Summarization using key phrases, in 'Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing', 1998.

[11] E. Cambouropoulos, Melodic cue abstraction, similarity, and category formation: A formal model. Music Perception 18 (3): 347–70, 2001.

[12] P.Toiviainen and T. Eerola, "A method for comparative analysis of folk musci based on musical feature extraction and neural networks", International Symposium on Systematic and Comparative Musicology III International Conference on Cognitive Musicology",2001.

[13] R. B.Dannenberg, "Music representation issues, techniques, and systems" Computer Music Journal 17 (3): 20–30, 1993.

[14] R. B.Dannenberg , B. Thom, and D. Watson. 1997. A machine learning approach to musical style recognition. Proceedings of the International Computer Music Conference. 344–7.