ISSN 1512-1232

CLASSIFICATION OF NORTH INDIAN MUSICAL INSTRUMENTS USING SPECTRAL FEATURES

M. Kumari*, P. Kumar** and S. S. Solanki*

* Department of Electronics and Communication, Birla Institute of Technology, Mesra, Ranchi, India
** Department of Space Engineering and Rocketry, Birla Institute of Technology, Mesra, Ranchi, India

Abstract

Musical instrument recognition finds its application in the field of Musical Information Retrieval such as automatic tagging, cataloguing of music for search engine, automatic musical transcription, genre classification etc. The present work has been emphasized on the identification of different North Indian Musical Instruments such as Flute, Sitar, Dholak, Bhapang, and Mandar. The spectral features were obtained from framed music signal and in order to classify them with each instrument spectral features such as spectral centroid, Spectral Spread, Spectral Skewness, Spectral Kurtosis, Spectral Slope, Spectral Rolloff, are obtained from MFCC feature extraction. Auto Correlation was also obtained for better classification of the instruments. It has been found that MFCC with 13 coefficients gives a better classification of monophonic instruments (Indian musical instruments) than any other feature. However, it was observed that Auto Correlation with 12 coefficients classified the instruments reasonably well. This technique is being attempted for the first time in classifying North Indian musical instruments.

Keywords*: Musical instrument, feature extraction, spectral feature, MFCC, Autocorrelation*

1. INTRODUCTION

In the present world, much effort is required for the tagging and cataloguing of music signals if it is done manually. It requires a trained musician or a prior knowledge of recording of the signals. The classification of such musical signal also helps in the mixing and demixing of Music researches are being understands into both single instrument (monophonic) and multi instrument (polyphonic) classification techniques. While polyphonic classification is a more general problem, monophonic classification still have significant relevance as polyphonic signal is a combination of monophonic signals.

The first identified work on classifying musical instruments was by Brown in 1997 [1]. This was followed on by work in 1998 [2]–[4] and 1999 [5] where Brown used cepstral coefficients based on the constant-Q transform and a k-means classifier to differentiate between recordings of oboe and saxophone with an error rate of 15%. In 2000, Eronen [6] also used cepstrum coefficients, along with 21 other complementary features including spectral centroid, spectral spread, rise and decay time, frequency and amplitude modulation rate and width and fundamental frequency. Using 30 instruments playing a single note, the reported accuracy was 75–80%. Eronen later [7], [8] explored a wider range of feature vectors, including both mel-frequency [9] and linear prediction cepstral and delta coefficients. This work included an analysis of 23 features and how successful each feature was individually for classification. It was found that mel-frequency cepstrum coefficients (MFCCs) alone were able to correctly classify in 20–30% of cases using 29 instruments. Also in 2004, Livshin and Rodet [10], [11] identified 20 important features for real time solo classification. Using these 20 features, they reported solo classifications of 85% using a k-Nearest Neighbour algorithm (compared to only 88% when using 62 different spectral features from

ISSN 1512-1232

the CUIDADO project [12]). These 20 features also performed reasonably well in duet recognition. In 2006, Essid et al. [13] used instrument hierarchies that were inspired by classical instrument groupings, and then inferred by automatic clustering, in an experiment to improve recognition results using SVMs. 540 signal processing features were considered in this study, from which a set were chosen using automatic feature selection [14]. Essid et al. [15] then went on to examine 150 feature vectors which they chose using inertia ratio maximisation and genetic algorithms, again using GMMs and SVMs.

The outline of the paper is the following. The present work focuses on the identification of five different north Indian musical instruments namely Sitar, Flute, Dholak, Bhapang and Mandar. We first present the set of signal processing features used and propose new features that prove to be useful for instrument recognition. The feature selection strategy as well as the classification technique is then described. Finally, we proceed to the experimental study.

2. Feature Extraction:

Many features have been proposed for musical instrument recognition describing various sound qualities. A number of these features become quite hard to extract robustly when dealing with musical phrases. A block diagram in Fig. 1 shows the extraction of features from the instrumental signals. Thus, a set of features which can be extracted in a more or less straightforward manner was chosen. In the following, we present a brief description of the features used. All of them are extracted on a frame basis. Each frame was cut into frames of 25 milliseconds. These features can be split into three main groups: spectral features (statistical data derived from the frequency spectrum); perceptual spectral features (features derived from a Mel-style spectrum); and other features that don't fit into the first two categories (e.g. MFCCs and the autocorrelation).The features extracted are as follows:

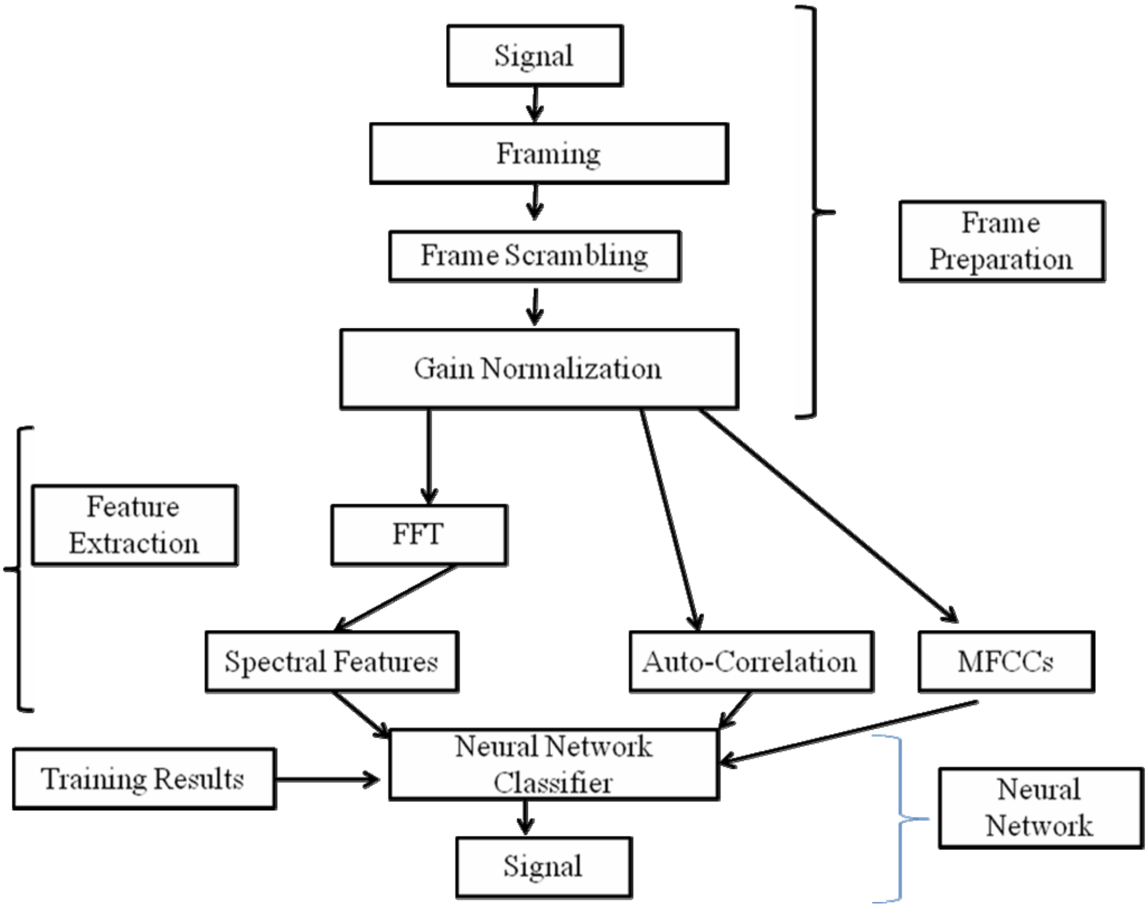


Figure 1: Block diagram of proposed work

ISSN 1512-1232

2.1 Spectral features:

Spectral features are single valued features, calculated using the frequency spectrum of a signal. Thus, for the time-domain signal x(t):

$$A(f) = |F[x(t)]|$$

2.1.1 Spectral Centroid:

The spectral centroid (μ) is the barycentre of the spectrum. It is a weighted average of the probability to observe the normalised amplitude [12].

Given A(f):

$$\mu = \int f.p(f)\delta f \qquad p(f) = \frac{A(f)}{\sum_f A(f)}$$

2.1.2 Spectral Spread:

The spectral spread (σ) is a measure of variance (or spread) of the spectrum around the mean value μ calculated in equation.

$$\sigma^2 = \int (f-\mu)^2.p(f)\delta f$$

2.1.3 Spectral Skewness:

The skewness is a measure of the asymmetry of the distribution around the mean value μ. The skewness is calculated from the 3rd order moment, m_3 [12].

$$m_3 = \int (f-\mu)^3.p(f)\delta f$$

2.1.4 Spectral Kurtosis:

Spectral kurtosis indicates the flatness or peakedness of the energy distribution. It is calculated from the 4th order moment, m4, using the value of μ [12]:

$$m_4 = \int (f-\mu)^4.p(f)\delta f$$

2.1.5 Spectral Slope:

The spectral slope (m) gives an indication of the rate of decrease of the amplitude A(f). The slope is simply a linear regression of the spectral amplitude [12].

$$m = \frac{1}{A(f)} \frac{N\sum_f A(f) - \sum_f f \times \sum_f A(f)}{\sum_f f^2 - \left(\sum_f f\right)^2}$$

2.1.6 Spectral Roll off:

The spectral roll off point (f_c) is the frequency for which 95% of the signal energy is below this frequency.

$$\sum_0^{f_c} A^2 f = 0.95 \sum_0^{f_{ny}} A^2(f)$$

Where f_{ny} is the Nyquist frequency.

ISSN 1512-1232

2.2 Mel Frequency Cepstral Coefficients:

Mel Frequency Cepstral Coefficients (MFCCs) are cepstral coefficients used for representing audio in a way that mimics the physiological properties of the human auditory system. MFCCs are commonly used in speech recognition and are finding increased use in music information recognition and genre classification systems. The cepstrum of a signal is the Fourier transform of the logarithm (decibel) signal (with unwrapped phase) of the Fourier transform of a signal. In the Mel frequency cepstrum, the frequencies are scaled logarithmically using the Mel scale. A block diagram in fig. 2 shows the process of feature extraction of MFCC.

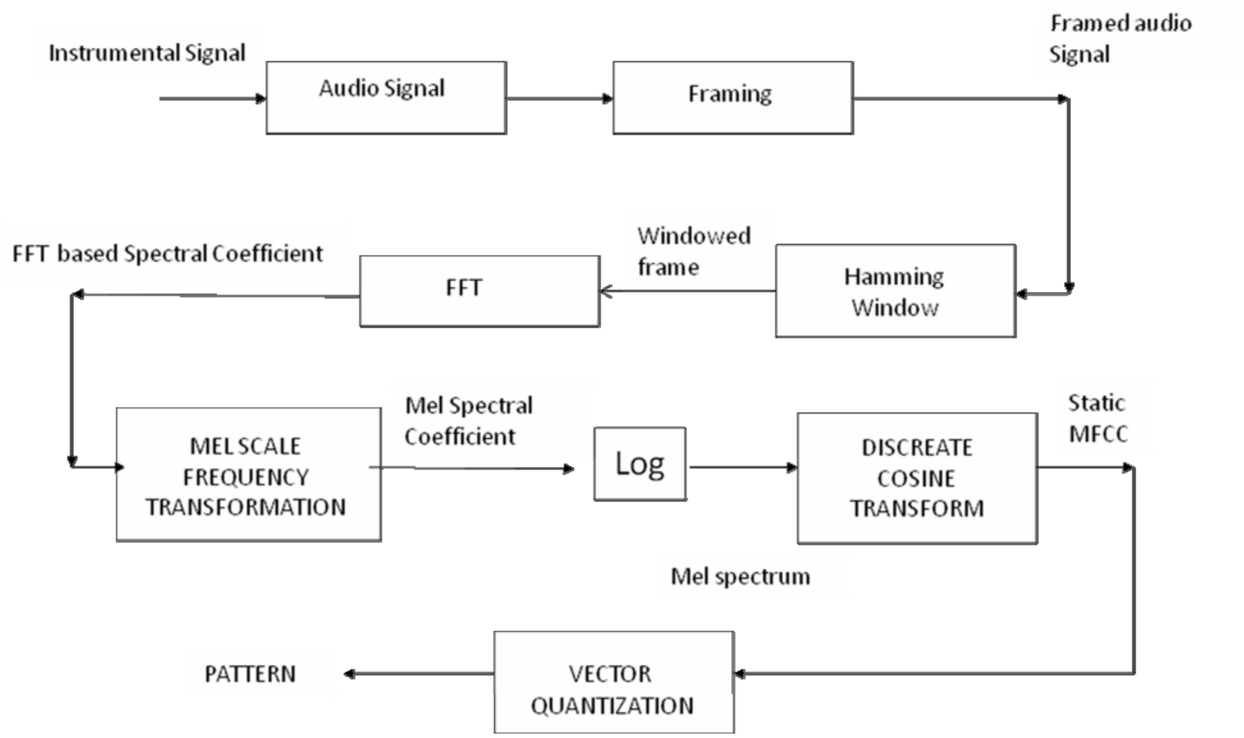


Figure 2: Block diagram of MFCC

2.3 Autocorrelation:

The autocorrelation of a signal is a measure of how well a signal matches with a time shifted version of itself. The autocorrelation of a frame represents the distribution of the signal spectrum but in the time domain. This feature was demonstrated to provide a good descriptor for classification by Brown [1].

3. Experimental procedure:

Five different instrumental signals were selected and the sound samples are digitized at a sampling rate of 44100/sec (16 bit per sample) and stored as wave files. Amplitude and frequency time varying curves of partials were measured based on Fast Fourier Transformation. For this the data was first segregated into frames of 25 milliseconds. Then the frames have been grouped in random order so as to remove the rhythm and tonality. The frames with silence were rejected. The above features are being compared and also the sound qualities of those instruments have been compared.

ISSN 1512-1232

4. Classification:

Artificial Neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes. A multilayer, feed forward neural network with one hidden layer was used for classification. A feed-forward neural network is one in which the neurons do not form a directed cycle. That is, a neuron in layer i − 1 is connected to every neuron in layer i, but to no other neurons in layer i − 1. This network is typically arranged into an input layer, one or more hidden layers and an output layer. Quasi – Newton Optimisation has been implemented for this purpose. First one third of the frames were taken for training and rest was taken for testing purpose.

Quasi Newton's method is based on a quadratic model $\tilde{E}$(w) of the error function E(**w**) and uses the first three terms in a Taylor series expansion of E about the current weight vector **w.**

$$\tilde{E}(w+\Delta w) = E(w) + \nabla E(w)^T \Delta w + \frac{1}{2}\Delta w^T \nabla^2 E(w) \Delta w$$

This is a quadratic function that is minimised by solving $\nabla\tilde{E}(w+\Delta w) = 0$, leading to Newton's equation:

$$\Delta w = -[\Delta^2 E(w)]^{-1} \nabla E(w) = -[H(w)]^{-1} \nabla E(w)$$

Where **H** is the Hessian matrix with components

$$H_{ij} = \frac{\partial^2 E}{\partial w_i \partial w_j}$$

5. Result and discussion:

Present work constitutes the identification of different Indian musical instruments using various spectral features. Several features such as spectral centroid, spectral slop, spectral kurtosis, spectral rolloff and MFCC were calculated for each instrument in order to identify the best suited feature for instrument classification. Finally six signals of sitar were selected so as to validate the features of each signal with other. The data obtained for minimum, maximum, mean, and standard deviation has been tabulated for each of the features of different sitar samples in table 1-6. It can be observed from the tables that the minimum, maximum, mean, and standard deviation, are almost similar for all the features of different sitar samples however slight variation could be observed for the case of spectral kurtosis. Histogram have been plotted for each features for better understanding of features, In order to have the more detailed analysis MFCC were obtained for 13 coefficients which has been shown in figure 5 which is almost similar for all the sitar files. The analysis shows that the features for the same instruments are almost similar; however more detailed analysis is required for the case of instruments played in different ways.

Then after, the analysis were carried out for the one of five different instruments namely Flute, Sitar, Dholak, Bhapang, and Mandar. The data obtained for minimum, maximum, mean, and standard deviation has been tabulated in table 7-12. The histogram plotted for each of different instruments. However better differentiation of instruments was seen for the case of spectral centroid and spectral rolloff. MFCC were calculated for each instruments has been shown in the Figure 6 It can be obtained from the fig that a large amount of variation could be captured for the case of each of the instruments. However, the pattern obtained for the instruments of same family e.g. Dholak, Bhapang, Mandar was almost similar. Autocorrelation was obtained for different instruments which have been shown in Fig. 7. Confusion matrix was obtained using 11, 12, and 13 coefficients of MFCCs which has been presented in Table 13-15. It can be obtained from the data the confusion matrix using 13 coefficients was the most suitable for the differentiation of instruments. Confusion

ISSN 1512-1232

Matrix was also obtained using 11, 12 and 13 coefficients of Autocorrelation which has been presented in Table 16-18.

6. Conclusions:

Spectral features were obtained for 5 different instrumental signals e.g. Flute, Sitar, Dholak, Bhapang, and Mandar. Spectral centroid and spectral rolloff shows better classification of Indian instruments. MFCC were obtained for each of the instruments. It can be concluded that confusion matrix obtained using 13 coefficients of MFCC shows better classification of different instruments. Confusion Matrix was also obtained using 11, 12 and 13 coefficients of Autocorrelation where it was observed that the confusion matrix obtained using 12 coefficients preformed reasonably well; however, the results obtained for the classification using 12 coefficients of Autocorrelation are much better than MFCC.

References:

1. J. Brown, Cluster-based probability model for musical instrument identification Journal of the Acoustical Society of America, 101, p. 3167, 1997.
2. Musical instrument identification using autocorrelation coefficients, in Proceedings of the International Symposium on Musical Acoustics, 1998.
3. Computer identification of wind instruments using cepstral coefficients, in Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, 1998.
4. J. Brown, Computer identification of wind instruments using cepstral coefficients, Journal of the Acoustical Society of America, 103, p. 2967, 1998.
5. J. C. Brown, Computer identification of musical instruments using pattern recognition with cepstral coefficients as features, The Journal of the Acoustical Society of America, 105, 3, p. 1933, 1999.
6. A. Eronen and A. Klapuri, Musical instrument recognition using cepstral coefficients and temporal features, in IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, 2000, p. 753.
7. A. Eronen, Comparison of features for musical instrument recognition, in IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, 2001, pp. 19–22.
8. "Automatic musical instrument recognition," Master's thesis, Tampere University of Technology, 2001.
9. S. Davis and P. Mermelstein, Comparison of parametric representations for mono-syllabic word recognition in continuously spoken sentences, IEEE Transactions on Acoustics, Speech and Signal Processing, 28, no. 4, pp. 357 – 366, 1980.
10. A. A. Livshin and X. Rodet, Instrument recognition beyond separate notes - indexing continuous recordings, in Proc. Int. Computer Music Conf., 2004.
11. Musical instrument identification in continuous recordings, in Proc. of the 7th Int. Conference on Digital Effects, 2004, pp. 222 – 227.
12. G. Peeters, A large set of audio features for sound description (similarity and classification) in the cuidado project, 2004. [Online]. Available: http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters 2003 cuidadoaudiofeatures.pdf.
13. S. Essid, G. Richard, and B. David, Hierarchical classification of musical instruments on solo recordings, in IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, 2006, pp.817.
14. I. Guyon and A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research, 3, pp. 1157–1182, 2003.
15. S. Essid, G. Richard, and B. David, Musical instrument recognition by pairwise classification strategies, IEEE Transactions on Audio Speech and Language Processing, 14, no. 4, p. 1401, 2006.

ISSN 1512-1232

Table: 1

Spectral centroid	Minimum	Maximum	Mean	Stddev
Sitar1	28.2231	103.25	56.9559	13.4255
Sitar2	24.2425	96.8062	55.3376	12.8351
Sitar3	28.5422	90.6378	55.0702	12.2266
Sitar4	23.6875	88.741	53.7504	13.0358
Sitar5	26.6301	88.0786	53.595	12.0776
Sitar6	25.2209	88.4905	52.0501	11.0189

Table: 2

Spectral spread	Minimum	Maximum	Mean	Stddev
Sitar1	32.8362	114.417	73.5569	17.8209
Sitar2	33.8934	116.5	73.3407	18.3714
Sitar3	33.6733	117.603	75.9048	19.0257
Sitar4	32.4348	122.127	73.3544	19.4133
Sitar5	33.2966	111.027	71.1414	18.056
Sitar6	33.66	113.497	70.3813	18.5825

Table: 3

Spectral skewness	Minimum	Maximum	Mean	Stddev
Sitar1	1.46444	6.85465	3.27773	0.924094
Sitar2	1.76167	6.77934	3.30324	0.852483
Sitar3	1.8478	6.83999	3.21862	0.766556
Sitar4	1.71882	7.35162	3.41748	0.973938
Sitar5	1.85684	6.70769	3.40265	0.874314
Sitar6	1.79968	6.0066	3.45118	0.748727

Table: 4

Spectral Kurtosis	Minimum	Maximum	Mean	StdDev
Sitar1	5.56318	73.9721	17.6053	10.0441
Sitar2	6.17783	67.8684	17.7657	9.6622
Sitar3	5.92081	68.8769	16.6592	8.49949
Sitar4	5.4513	69.0929	18.8559	11.2907
Sitar5	6.56383	67.7339	18.8269	10.0302
Sitar6	6.51026	62.0232	19.1453	8.94012

Table: 5

Spectral Slope	Minimum	Maximum	Mean	StdDev
Sitar1	-4.72E-07	-3.16E-07	-4.12E-07	2.79E-08
Sitar2	-4.80E-07	-3.29E-07	-4.16E-07	2.66E-08
Sitar3	-4.71E-07	-3.42E-07	-4.16E-07	2.54E-08
Sitar4	-4.81E-07	-3.46E-07	-4.19E-07	2.71E-08
Sitar5	-4.75E-07	-3.48E-07	-4.19E-07	2.51E-08
Sitar6	-4.78E-07	-3.47E-07	-4.22E-07	2.29E-08

ISSN 1512-1232

Table: 6

Spectral Rolloff	Minimum	Maximum	Mean	StdDev
Sitar1	1162.79	7192.09	3059.3	1081.81
Sitar2	1248.93	5512.5	2940.23	917.108
Sitar3	516.797	5512.5	2696	933.278
Sitar4	818.262	6029.3	2821.6	955.976
Sitar5	559.863	6072.36	2929.02	944.387
Sitar6	732.129	4694.24	2680.76	770.644

Table: 7

Spectral Centroid	Minimum	Maximum	Mean	StdDev
Flute	31.2323	125.812	59.2048	18.1904
Dholak	12.0902	95.5157	52.1997	18.8119
Sitar	29.5383	88.6773	54.9348	12.5576
Bhapang	17.0738	90.4821	47.9595	16.7136
Mandar	11.8273	91.3164	46.0483	16.2012

Table: 8

Spectral Spread	Minimum	Maximum	Mean	StdDev
Flute	45.3969	154.501	91.579	21.5716
Dholak	24.2201	124.757	86.7577	24.5444
Sitar	33.6733	117.603	76.0308	19.0924
Bhapang	22.7765	125.469	84.2957	23.5486
Mandar	24.6687	120.7	84.6379	23.6707

Table: 9

Spectral Skewness	Minimum	Maximum	Mean	StdDev
Flute	1.04342	6.04962	3.25745	1.11372
Dholak	1.56254	10.1023	3.56393	1.69955
Sitar	1.92542	6.83999	3.20138	0.791268
Bhapang	1.75953	10.2344	3.76228	1.64572
Mandar	1.86167	10.3102	3.7928	1.64371

Table: 10

Spectral Kurtosis	Minimum	Maximum	Mean	StdDev
Flute	2.67276	46.6245	15.034	9.44108
Dholak	4.50003	148.105	20.2062	22.6825
Sitar	5.92081	68.8769	16.5409	8.90168
Bhapang	5.0555	155.16	21.4066	22.1734
Mandar	5.56423	146.884	21.4962	21.1877

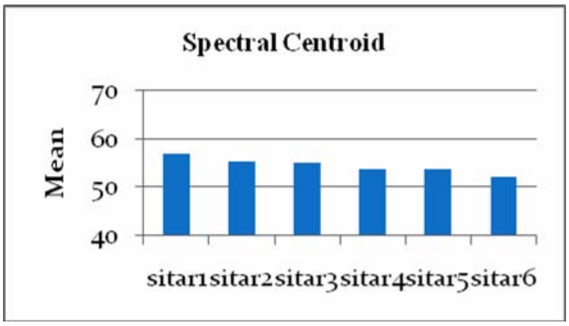
Table: 11

Spectral Slope	Minimum	Maximum	Mean	StdDev
Flute	-4.66E-07	-2.69E-07	-4.08E-07	3.78E-08
Dholak	-5.05E-07	-3.32E-07	-4.22E-07	3.91E-08
Sitar	-4.69E-07	-3.46E-07	-4.16E-07	2.61E-08
Bhapang	-4.95E-07	-3.43E-07	-4.31E-07	3.47E-08
Mandar	-5.06E-07	-3.41E-07	-4.35E-07	3.36E-08

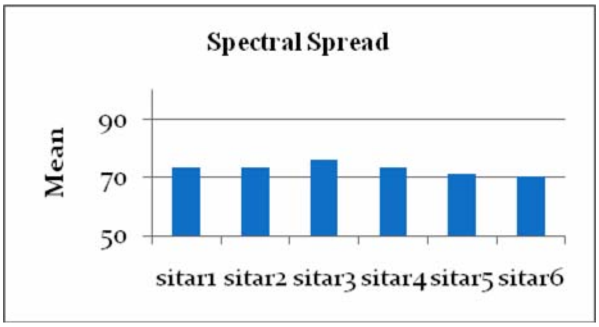
ISSN 1512-1232

Table: 12

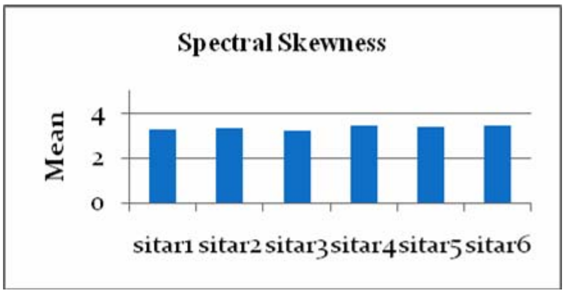
Spectral Rolloff	Minimum	Maximum	Mean	StdDev
Flute	645.996	2411.72	1398.05	215.302
Dholak	172.266	2670.12	1055.98	351.379
Sitar	516.797	5512.5	2594.04	937.485
Bhapang	387.598	2196.39	894.339	263.814
Mandar	172.266	2799.32	890.166	319.161



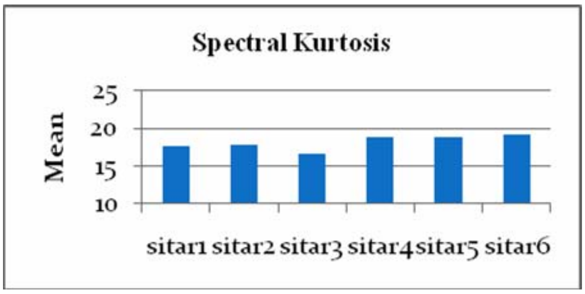
(a)



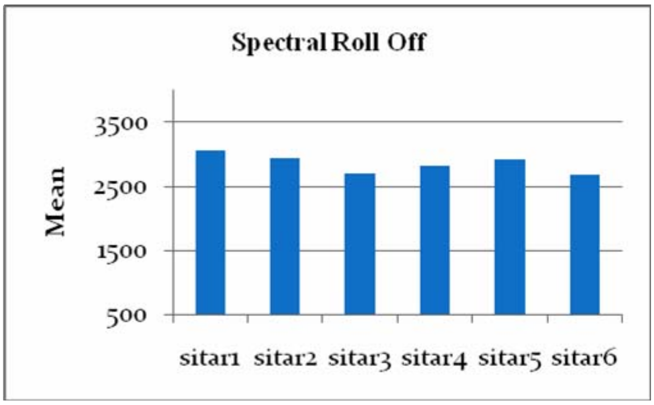
(b)



(c)



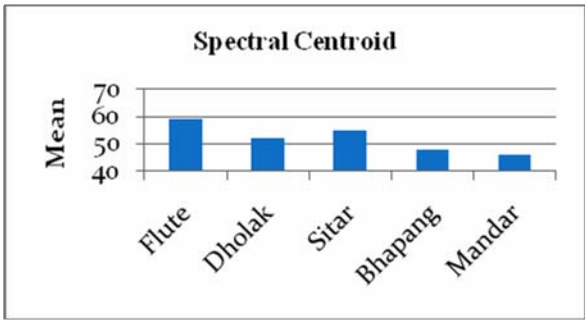
(d)



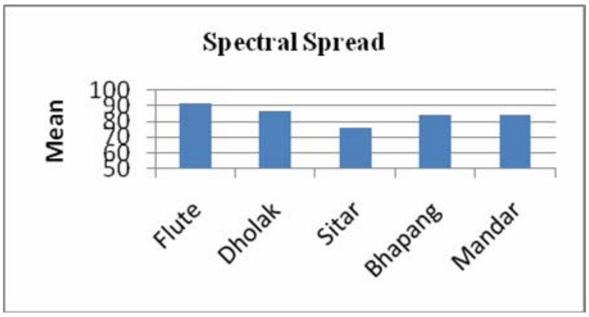
(e)

Figure 3(a-e): Spectral Feature of different Sitar samples

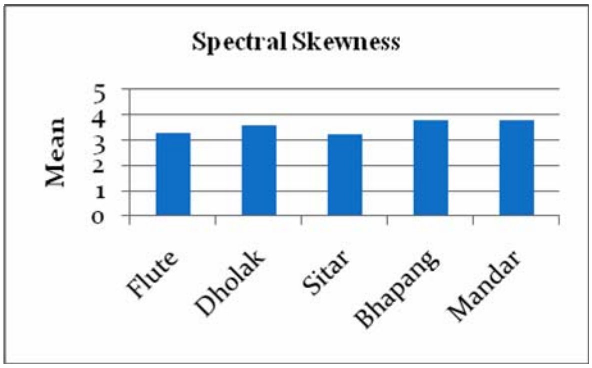
ISSN 1512-1232



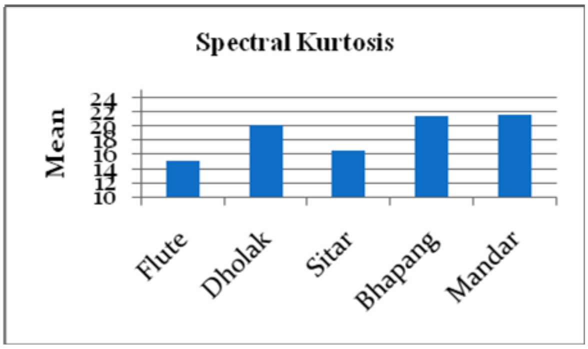
(a)



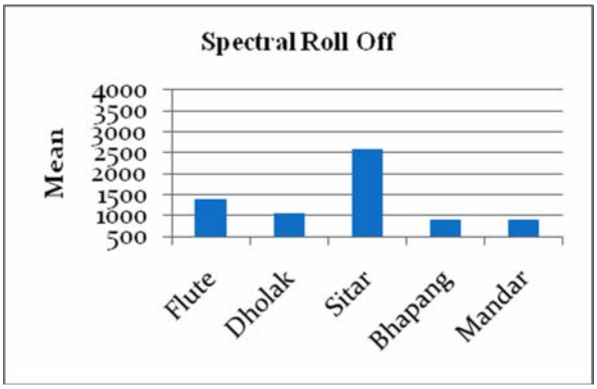
(b)



(c)



(d)



(e)

Figure 4(a-e): Spectral Feature of different Indian Musical Instrument

ISSN 1512-1232

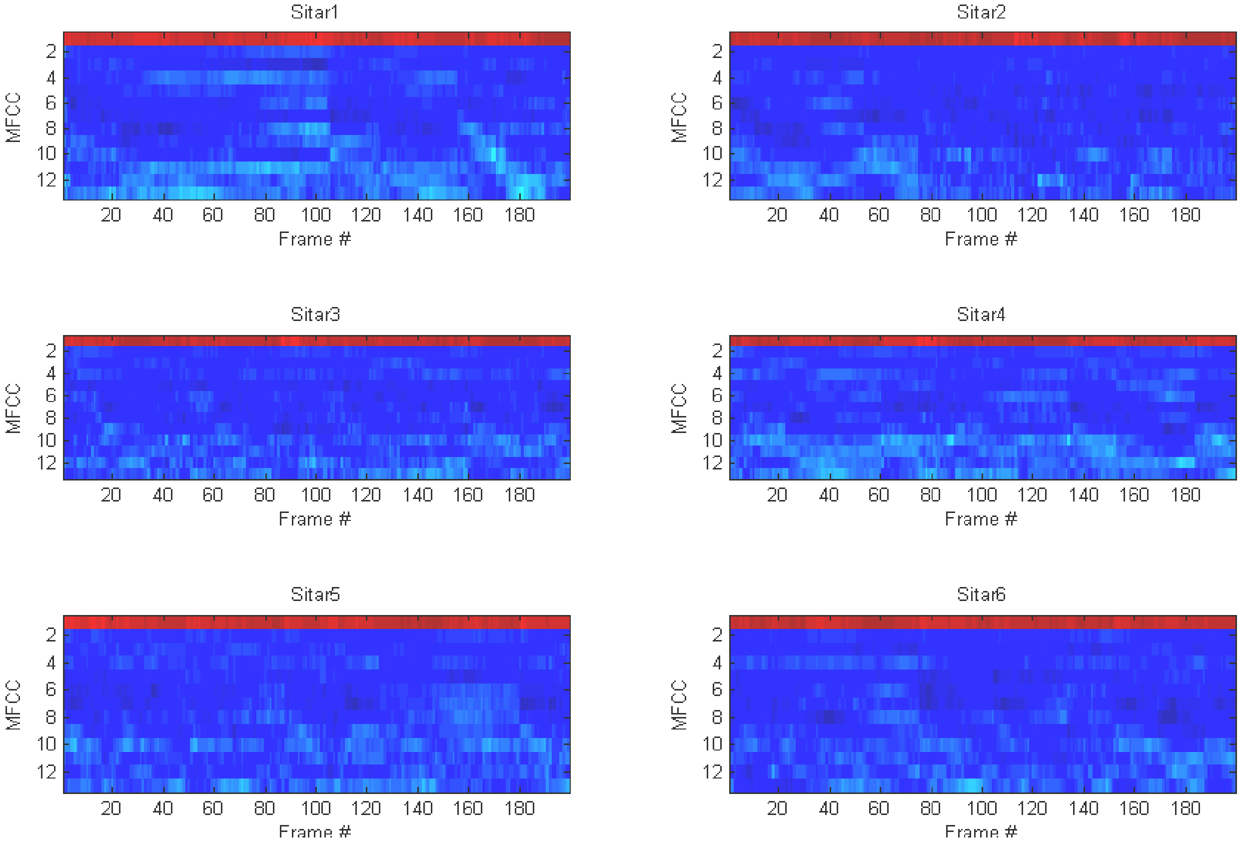


Figure.5: MFCC of different sitar sample

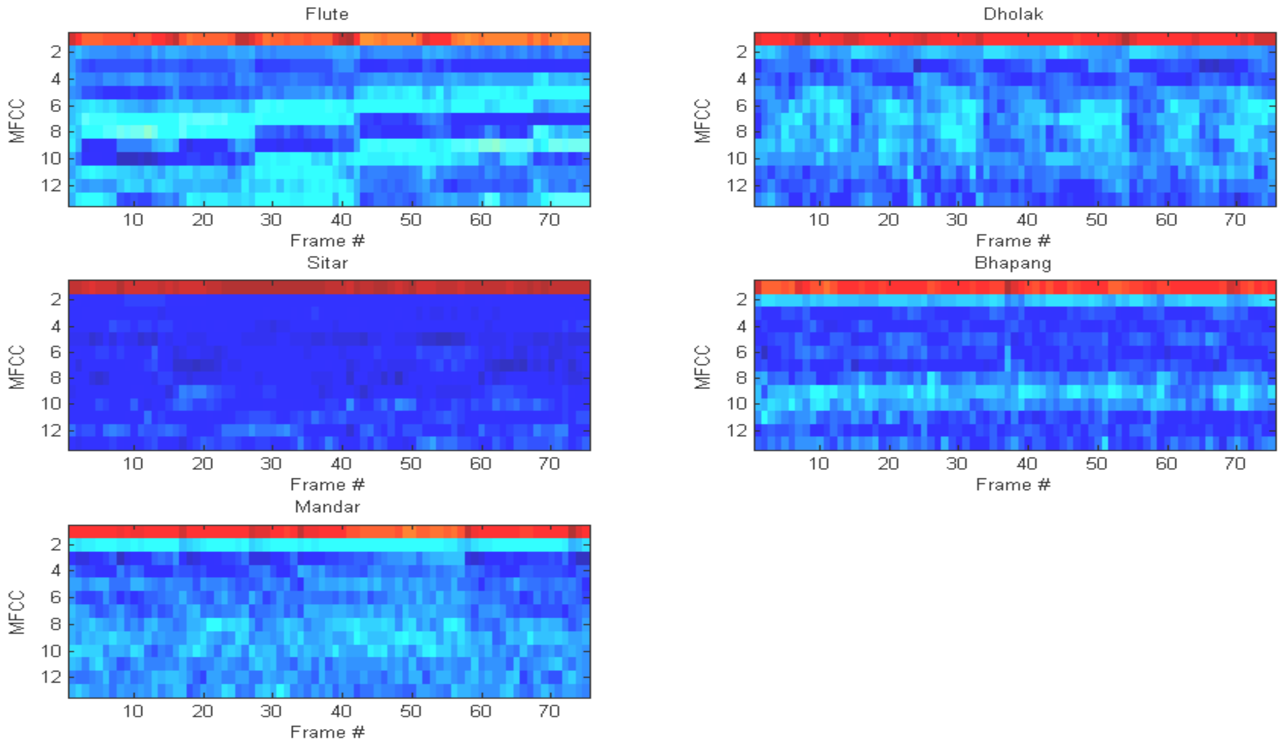


Fig.6: MFCC of Flute, Dholak, Sitar, Bhapang, Mandar

ISSN 1512-1232

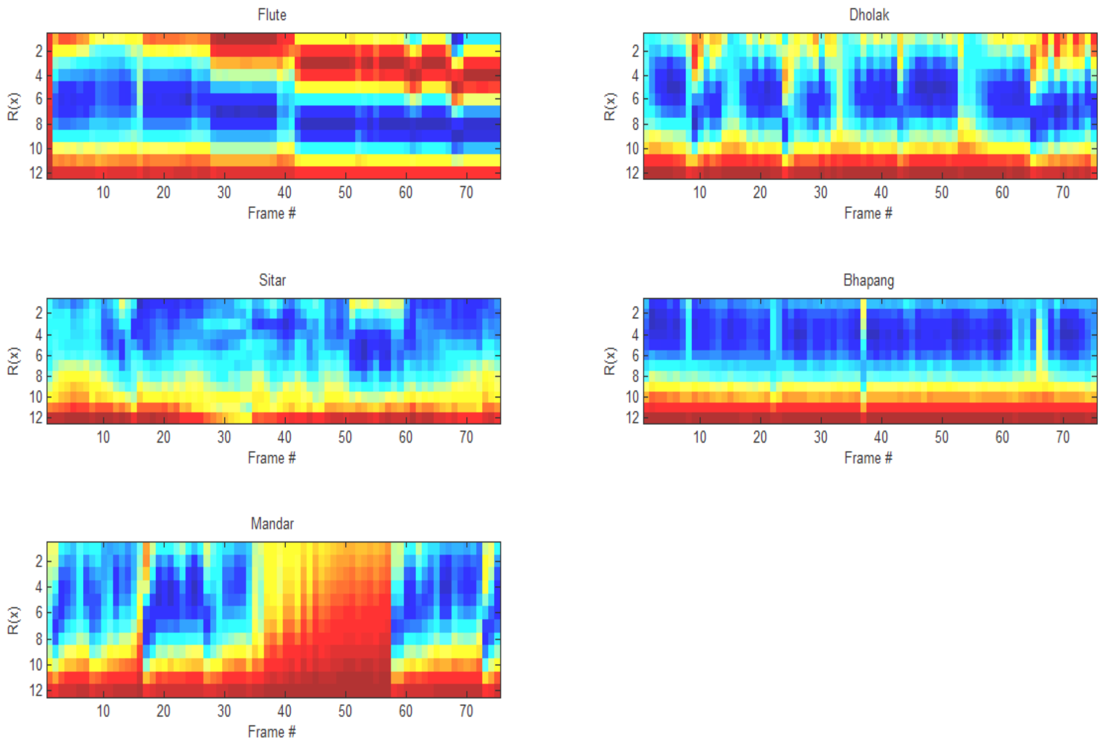


Fig.7: Autocorrelation of Flute, Dholak, Sitar, Bhapang, Mandar

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	**58.00%**	0.00%	0.00%	42.00%	0.00%
Dholak	4.70%	**66.70%**	5.30%	0.00%	23.30%
Sitar	0.70%	0.00%	**60.70%**	38.00%	0.70%
Bhapang	11.30%	3.30%	0.00%	**68.10%**	17.30%
Mandar	10.00%	0.70%	0.0%	5.30%	**84.00%**

Table: 13 Confusion matrix using MFCC(with 11 coefficients)

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	**61.00%**	37.00%	0.00%	0.00%	2.00%
Dholak	21.00%	**77.00%**	0.00%	0.00%	2.00%
Sitar	40.00%	0.00%	**60.00%**	0.00%	0.00%
Bhapang	5.30%	28.70%	0.00%	**64.70%**	1.30%
Mandar	0.00%	11.30%	5.3%	16.70%	**66.70%**

Table: 14 Confusion matrix using MFCC(with 12 coefficients)

ISSN 1512-1232

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	**64.0%**	0.0%	12.0%	24.0%	0.00%
Dholak	0.00%	**63.30%**	0.0%	26.7%	10.0%
Sitar	23.30%	6.70%	**70.00%**	0.00%	0.00%
Bhapang	8.70%	2.00%	0.00%	**89.30%**	0.00%
Mandar	8.70%	0.70%	0.0%	0.00%	**90.70%**

Table: 15 Confusion matrix using MFCC(with 13 coefficients)

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	**65.00%**	1.00%	30.00%	3.00%	1.00%
Dholak	9.30%	**80.70%**	1.30%	2.70%	6.00%
Sitar	0.70%	4.70%	**77.30%**	9.30%	8.00%
Bhapang	40.00%	0.00%	0.00%	**50.00%**	10.00%
Mandar	4.70%	0.70%	19.30%	20.70%	**54.70%**

Table: 16 Confusion matrix using Autocorrelation (with 11 coefficients)

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	**79.30%**	1.30%	11.30%	0.70%	7.30%
Dholak	6.00%	**70.70%**	1.70%	11.30%	5.30%
Sitar	2.00%	0.00%	**96.00%**	2.00%	0.00%
Bhapang	0.70%	2.00%	11.30%	**84.00%**	2.00%
Mandar	3.30%	7.30%	0.70%	5.30%	**83.30%**

Table: 17 Confusion matrix using Autocorrelation (with 12 coefficients)

ISSN 1512-1232

	Flute	Dholak	Sitar	Bhapang	Mandar
Flute	68.70%	10.30%	0.70%	19.70%	0.70%
Dholak	9.30%	80.70%	1.30%	2.70%	6.00%
Sitar	0.70%	4.70%	77.30%	9.30%	8.00%
Bhapang	40.00%	0.00%	0.00%	50.00%	10.00%
Mandar	4.70%	0.70%	19.30%	20.70%	54.70%

Table: 18 Confusion matrix using Autocorrelation (with 13 coefficients)

Article received: 2010-04-15