IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 01-07
www.iosrjournals.org

Indian Musical Instrument Recognition based MFCC Feature Set

Sankaye, S.R.[1], Tandon U.S.[2]
[1](MGM Dr.G.Y.Pathrikar College of Computer Science & Information Technology, Aurangabad, India)
[2](Department of Physics, College of Natural and Computational Science, Haramaya University Ethiopia)

Abstract: *As the music databases grow in size and number, the retrieval of music information is becoming an important task for various applications. There has been shift with the researcher's trends from speech signal processing to the musical information retrieval (MIR). In this study we endeavored for the recognition of Indian Musical Instruments sound sample recorded in natural environment. The Features considered for the recognition include zero crossing, root means Square Energy and Mel-Frequency Cepstral Coefficient. We achieved an aggregate of 77.5% correct results for the recognition of four instruments considered with highest value for the Harmonium at 90.00%.*

Keywords: *Avanadha Vadya, Ghana Vadya, Indian Musical Instrument, Mel-Frequency Cepstral Coefficient, RootMeans Square Energy, Sushir Vadya, Tantu Vadya, Zero Crossing*

I. Introduction

Automatic musical instrument recognition is a crucial task in solving difficult problems and also to provide useful information in sound source recognition areas, such as speaker recognition. Through the construction of computer systems that "listen", we may gain some new insights into human perception. This work describes the construction and evaluation of a musical instrument recognition system that is able to recognize the Indian musical instruments.

Musical instrument recognition is related to many other fields of research. The methods used in implementing musical instrument recognition systems are drawn from different technical areas. The preprocessing and feature extraction techniques can be taken from speech and speaker recognition. Commonly, classification is performed with statistical pattern recognition techniques.

Various attempts have been made to construct automatic musical instrument recognition systems. Researchers have used different approaches and scopes, achieving different performances. Most systems have operated on isolated notes, often taken from the same, single source, and having notes over a very small pitch range. The most recent systems have operated on solo music taken from commercial recordings. Polyphonic recognition [1] [2] has also received some attempts, although the number of instruments has still been very limited. The studies using isolated tones and monophonic [3] phrases are the most relevant in our scope.

A musical sound is said to have four perceptual attributes [4]: pitch, loudness, rhythms and timbre. These four attributes make it possible for a listener to distinguish musical sounds from each other. Pitch, loudness and duration are better understood than timbre and they have clear physical counterparts. For musical sounds, pitch is well defined and is almost equal to the fundamental frequency. The physical counterpart of loudness is intensity, which is proportional to the square of the amplitude of the acoustic pressure.

Pitch: It is purely a psychological term that relates to the actual frequency of a particular tone and relative position in the musical scale. It is the only attribute that varies in the first seven notes, while the rhythms stay the same.

Loudness: It is another psychological attribute that relates to how much energy an instrument creates. Loudness is the term refers to envelop, and it is used to describe how the loudness of sound changes over time.

Rhythms: It refers to the durations of a series of note, and to the way that they group together into units. Percussive sounds are the mostly used to convey the rhythmic dimension.

II. Research review

Analysis of music data and retrieval have become a very popular research field in recent years. The advances in signal processing and data mining techniques have led to intensive study on musical instrument detection and classification. Instrument detection techniques can have many potential applications. For instance, detecting and analyzing solo passages can lead to more knowledge about different musical styles and be further utilized to provide a basis for lectures in musicology. Various applications for audio editing, audio and video retrieval or transcription can be supported. Some of the prominent researcher work is enlisted.

Eronen [5] assessed the performance of mel-frequency cepstral coefficients (MFCC) features and spectral and temporal features such as amplitude envelope and spectral centroids for instrument classification.

Kitahara, Goto and Okuno [6] proposed the method for musical instrument identification based on F0-dependent multivariate normal distribution. The F0-dependent mean function represents the pitch dependency. His result showed the recognition rate at individual recognition rate up to 79.73%.

Emmanuel Vincent and Xavier Rodet [7] investigated the use of Independent Subspace Analysis (ISA) for instrument identification in musical recordings. They represented short-term log-power spectra of possibly polyphonic music as weighted non-linear combinations of typical note spectra plus background noise. These typical note spectra are learnt either on databases containing isolated notes or on solo recordings from different instruments. The model has some theoretical advantages over methods based on Gaussian Mixture Models (GMM) or on linear ISA.

Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler[8], discussed the methods based on the use of explicitly predefined signal features: the signal's amplitude envelope, spectral magnitudes and phases, time-frequency representations; and methods based on probabilistic signal models: model-based change point detection, surprise signals, etc. The experiments were performed on a database of commercial and noncommercial recordings covering a variety of musical styles and instrumentations. All signals were processed as monaural signals sampled at 44.1 kHz.

III. Experimental Work

In our study, to recognize the Indian musical instruments domain, a set of algorithmic procedure: preprocessing, feature extraction, training and recognition has been used as shown in figure below.

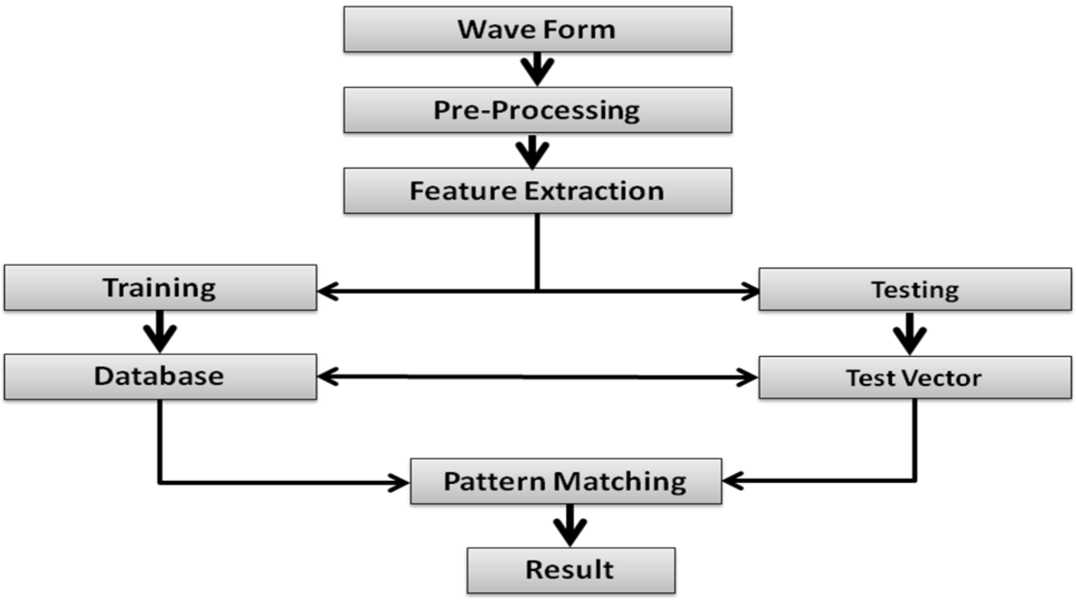

Figure 1: Phases of Musical Instrument Recognition System

The preprocessing stage consists of musical instrument sound recordings, editing, and storing musical sounds in a database. For the purpose of this study, several recording sessions were organized for different instruments by different performers. We did recording for only four instruments viz. Flute, Harmonium, Clarinet and Mandolin in the database. In this stage, the noise is removed from the input signal. Although the raw sound samples of these musical instruments were recorded in anechoic chamber still there exist some extra sound i.e. noise (unwanted sound signal) such as breathing of the musician, room noise etc. It was removed manually.

Recorded sounds were transferred through Channel Audio Mixer via the sound recording software Cool Edit Pro into the computer hard disc. The next step was the preparation of the database recorded sound. The sound samples were recorded at sampling rate of 44100Hz to get CD quality, 16-bit, stereo type. For the purpose of our study the Musical sound samples were transformed from the stereo type to mono type maintaining the CD quality of the sound samples i.e. 44.1 KHz.

Table 1: Musical Instrument Sound Database Recorded

Sr.No.	Name of Instrument	No. of Sample	No. of Performer	Note
1	Clarinet	50	2	Mend + Performance
2	Mandolin	50	2	Octave + mend + Performance
3	Harmonium	32	2	Mend + Performance
4	Flute	50	2	Mend + Performance

During the Feature Extraction step, knowledge base information on the musical instruments is built. The sounds cannot be identified very easily hence a number of features of the acoustical signals are needed. It results in the creation of feature vector. These sets of feature vector are characteristic for most musical instrument, depending upon which the decision process is carried out. Each musical instrument has its own signature or defining features. In the time domain, each musical sound can be represented by a complex wave graph. The attribute most visible in the time domain is the amplitude. For example in figure 2, a plot of the sound wave shows the differing amplitude for a Harmonium and Mandolin.

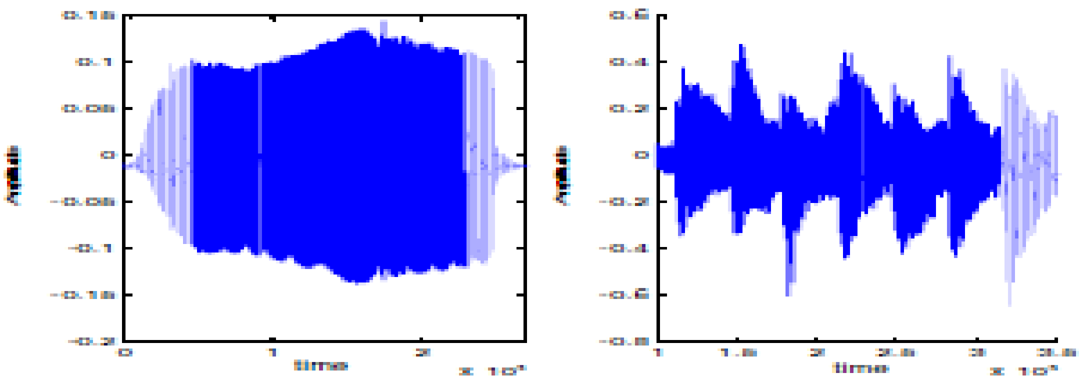

Figure 2: Waveform of the Musical Instrument Harmonium and Mandolin

For any musical tone, if a segment of the wave graph is enlarged we can see the periodic nature of the tone. For example, the Harmonium and Mandolin tone is illustrated in figure 3.

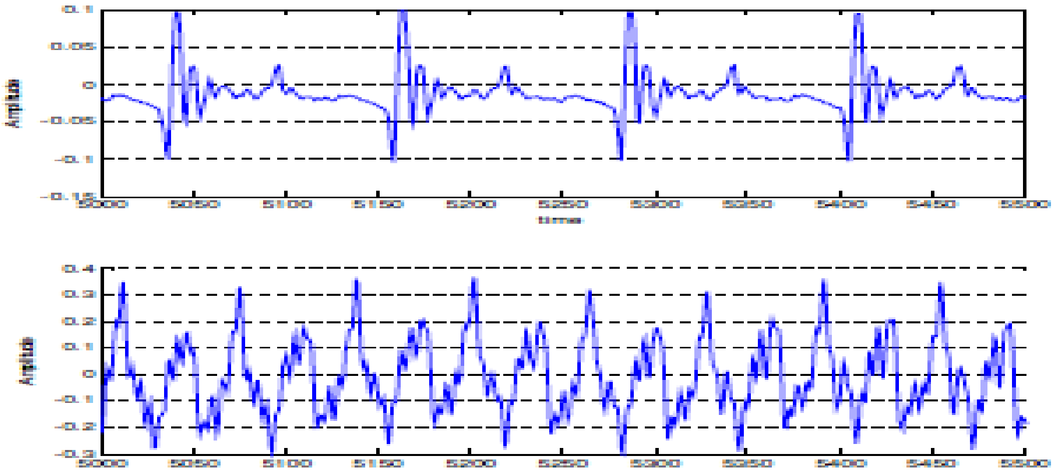

Figure 3: Periodic Nature of Musical Tone of Musical Instrument

The following are a perceptual feature that was used in our study:

1) **Zero-Crossing Rate (ZCR):** It is an indicator for the noisiness of the signal, often used in speech processing applications i.e. it is used to detect the noise in the signal. It relates information about the spectral content of the waveform. ZCR gives us discriminating periodic signals (for small ZCR) from noisy signals (for high ZCR) [9]. Figure 4 and 5 shows the Zero Crossing Rate for the Musical Instruments viz. Harmonium and Mandolin respectively.

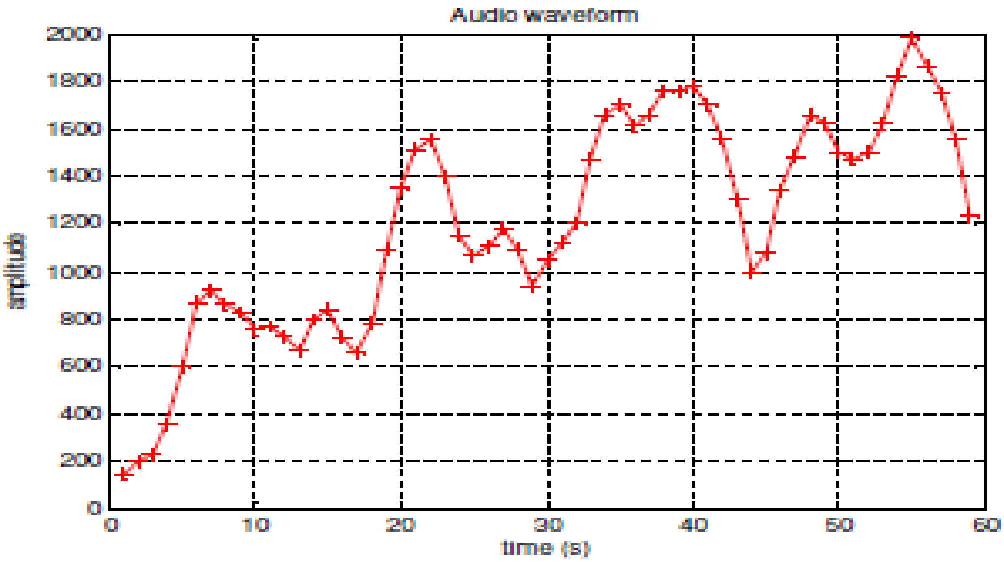

Figure 4: ZCR of the Musical Instrument *Harmonium*

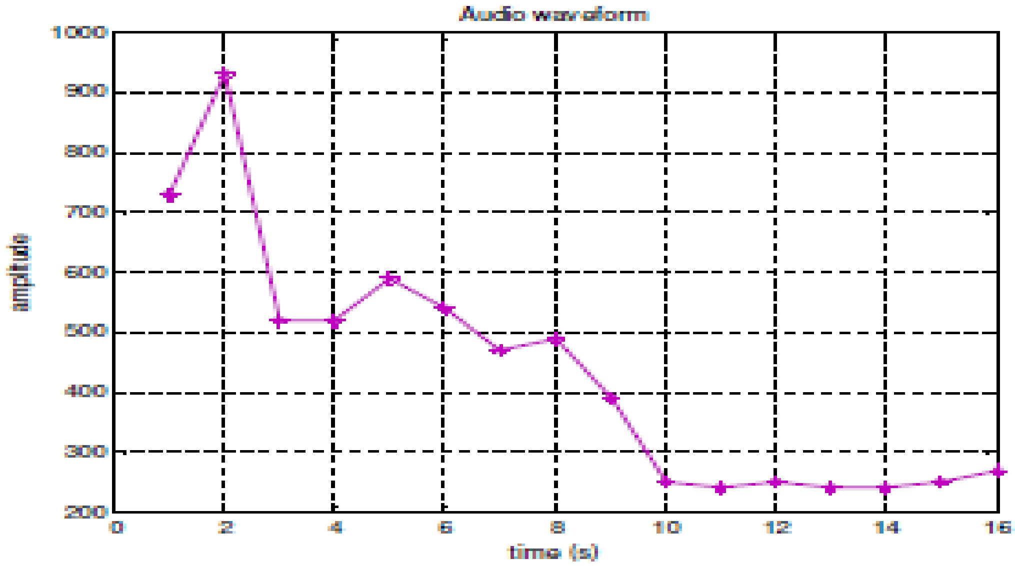

Figure 5: ZCR of the Musical Instrument *Mandolin*

2) **Root-mean-square (RMS):** RMS summarizes the energy distribution in each frame and channel over time. RMS is the global energy of the signal that is computed simply by taking the square root of the arithmetic mean of the squares of the amplitude. The RMS-energy envelope, on a linear scale, is also used to extract features measuring amplitude modulation (AM) properties like Strength, frequency, and heuristic strength (term used by Martin [10]). Table 2 shows the RMS Values obtained for all the four musical instruments for the sample 20 sound.

Table 2: Root-mean Square Energy of Musical Instruments.

Sample	**Indian Musical Instrument**			
	Clarinet	**Flute**	**Mandolin**	**Harmonium**
s1	0.0446	0.2218	0.0774	0.0253
s2	0.0528	0.1397	0.0410	0.0232
s3	0.0522	0.2683	0.0273	0.0222
s4	0.0811	0.2976	0.0322	0.0187
s5	0.0554	0.3024	0.0251	0.0173
s6	0.0872	0.2907	0.0255	0.0257
s7	0.0710	0.2640	0.0611	0.0325
s8	0.0421	0.4681	0.1212	0.0389

s9	0.0392	0.2314	0.0623	0.0288
s10	0.0238	0.2993	0.0520	0.0364
s11	0.0516	0.1826	0.0401	0.0493
s12	0.0438	0.2361	0.0202	0.0413
s13	0.0437	0.2891	0.0089	0.0398
s14	0.0179	0.1431	0.0119	0.0510
s15	0.0856	0.2509	0.0589	0.0334
s16	0.1165	0.2407	0.0589	0.0398
s17	0.1271	0.3004	0.0384	0.0510
s18	0.0947	0.3553	0.0467	0.0473
s19	0.1147	0.2444	0.1295	0.0640
s20	0.1520	0.4702	0.0696	0.0515

3) Mel-frequency cepstral coefficients (MFCCs): Mel-Frequency Cepstrum Coefficient [11] [12] is the frequency bands that are positioned logarithmically on the mel-scale which approximate the human auditory system response. Thus it gives better results with comparison to linearly spaced frequency band. Using Matlab Audio toolbox MFCC Feature set for first thirteen features were calculated using the modified MFCC for all four instruments sound signal nearly 50 samples of each. Table 3 shows the MFCC 13-feature vector set for the *Clarinet* Musical Instrument. We also computed Delta-MFCC as well as Delta-Delta MFCC features set for final feature Vector which consist of 34 Features based on MFCC.

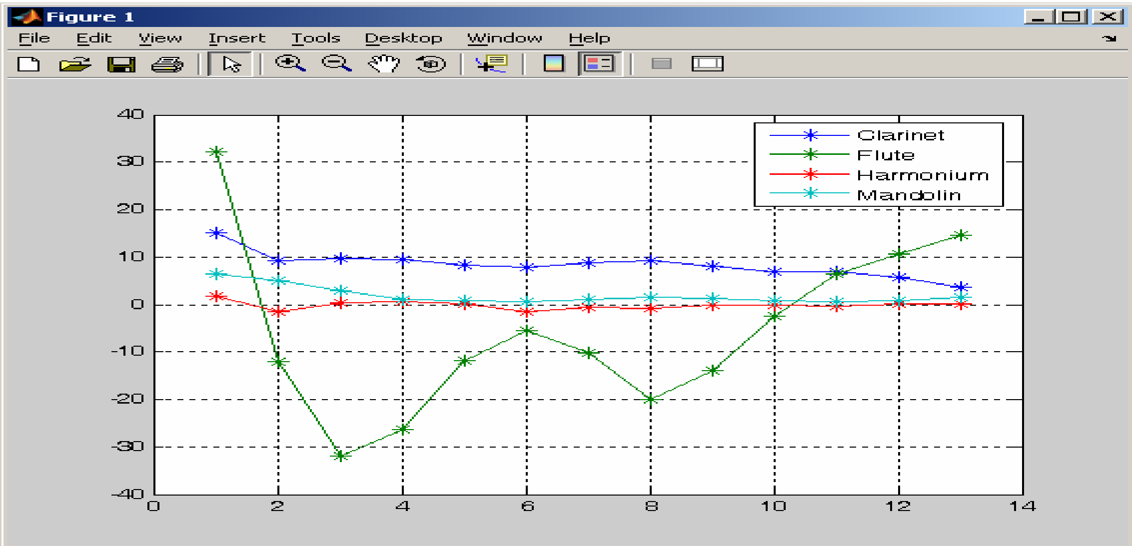

Figure 6: Mean of MFCC Feature Vector Set of All four Musical Instruments.

Once the data sheet for all the four instruments were recorded and calculated, we calculated the mean, median and standard deviation of individual instruments. Figure 6 above shows the Mean of all the fours musical instruments.

Table 3: Mel-frequency Cepstrum Coefficient of Clarinet Musical Instrument.

Sample	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
s1	6.23	7.71	6.96	6.42	6.34	6.23	6.29	6.30	6.42	6.69	7.04	7.21	7.11
s2	16.31	13.62	11.38	12.59	13.38	12.86	12.18	12.05	12.51	12.19	11.56	11.53	13.12
s3	11.45	7.32	5.64	4.13	4.67	7.34	8.12	7.25	6.08	5.91	7.89	9.39	10.90
s4	22.96	17.23	11.57	9.31	8.80	10.06	11.86	11.32	12.35	16.60	21.89	24.60	25.11
s5	19.59	13.61	7.23	5.80	4.80	5.70	9.31	10.82	14.75	18.73	20.58	19.03	15.49
s6	45.89	32.71	25.66	18.89	13.87	18.70	27.78	34.64	41.41	45.65	45.66	36.87	25.32
s7	37.30	17.99	19.55	15.08	10.45	19.16	29.07	36.12	38.33	30.74	17.37	5.12	-0.36
s8	13.86	1.69	5.13	6.14	3.08	6.53	13.81	14.86	8.61	3.60	-0.02	-5.97	-6.74

s9	11.45	-7.21	2.04	8.93	7.48	3.05	6.30	15.39	11.15	-2.97	-11.58	-8.71	-0.65
s10	7.33	1.72	3.38	4.76	2.95	3.22	6.14	6.20	0.43	-5.19	-5.16	-2.39	-0.25
s11	10.29	8.23	9.09	8.61	6.62	4.51	2.12	-1.03	-6.40	-9.32	-6.95	-6.14	-10.62
s12	9.08	4.22	7.67	8.59	3.48	0.35	-0.12	-2.46	-6.26	-7.30	-6.52	-8.47	-11.41
s13	9.06	4.21	7.64	8.57	3.49	0.34	-0.14	-2.46	-6.23	-7.29	-6.53	-8.45	-11.36
s14	8.25	7.98	7.40	3.39	-1.21	-4.03	-6.11	-8.67	-10.68	-11.36	-10.88	-8.93	-6.04
s15	7.64	7.49	6.51	7.08	7.47	6.61	6.78	7.05	7.22	7.03	6.81	6.98	7.59
s16	27.38	23.80	20.71	22.33	22.32	21.62	21.92	21.58	21.19	21.38	23.28	25.25	28.79
s17	13.49	16.26	13.83	12.61	12.27	11.47	11.58	11.87	12.74	14.27	15.83	16.90	16.85
s18	10.71	9.86	8.17	8.78	9.09	8.59	8.57	8.75	8.76	8.67	8.83	9.28	10.37
s19	12.25	11.69	10.38	10.35	10.20	10.14	9.83	9.52	9.80	10.03	10.97	12.11	13.93
s20	14.99	16.77	14.04	13.13	12.77	11.56	11.89	12.23	13.08	15.03	17.12	18.64	18.73

Based on the feature Vectors Set, we prepared a Model training which stored the feature vectors corresponding to the different classes of the Musical instruments as an input signal as a finite number of templates. In the testing step, the feature vector of the input signal is compared to the stored templates. Using the Euclidean distance, we were able to prepare the decision matrix for various instruments and final results obtained are shown in table 4.

Table 4: Recognition Rate of Musical Instrument.

Sr.No.	Instrument Name	Recognition Rate
1	Clarinet	65 %
2	Flute	85 %
3	Harmonium	90 %
4	Mandolin	70 %

IV. Conclusion

We have attempted to obtain a feature extractor which retains the discriminative information through fitness function for the Indian Musical instrument recognition criteria. Although this is not the foremost but it definitely is an approach for the optimization of feature extraction for recognizing Indian musical instruments.

Acknowledgements

We would like to accord our sincere thanks to Pt. Gopal Mishraji for his valuable support by playing the Musical Instruments for our study. Also we would like to put word of thanks to Mr. Vishwanath N. Oak for providing us the studio to record the sound samples. We wish to express our thanks to Dr. S.C. Mehrotra for useful discussions.

References:

[1]. Slim ESSID*, Ga¨el RICHARD and Bertrand DAVID "Instrument Recognition in Polyphonic MusicBased on Automatic Taxonomies"

[2]. Tetsuro Kitahara, Masataka Goto, Kazunori Komatani, Tetsuya Ogata and Hiroshi G. Okuno "Instrument Identification In Polyphonic Music: Feature Weighting With Mixed Sounds, Pitch-Dependent Timbre Modeling, And Use Of Musical Context", 2005.

[3]. S. Essid, G. Richard, and B. David, "Musical instrument recognition on solo performance," in Proc. Eur. Signal Processing Conf., pp. 1288–1292 , 2004.

[4]. Antti Eronen , "Automatic Musical Instrument Recognition", Master of Science Thesis, 2001.

[5]. A. Eronen, "Comparison of features for music instrument recognition" in Proc. of the IEEE Workshopon Applications of Signal Processing to Audio and Acoustics, pp. 19–22, 2001.

[6]. T. Kitahara, M. Goto, H. G. Okuno, "Musical Instrument identification Based on F0-dependent multivariate normal distribution.", Proc. of IEEE Int. conference on acoustic, speech and signal processing, vol. 5, pp 421-424, April,2003.

[7]. Emmanuel Vincent and Xavier Rodet, "Instrument Identification In Solo And Ensemble Music Using Independent Subspace Analysis".

[8]. Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler "A Tutorial on Onset Detection in Music Signals", IEEE Transactions On Speech And Audio Processing, Vol. 13,No. 5, Sep'2005.

[9]. Slim Essid, G.Richard, "Musical Instrument Recognition by pairwise classification Strategies", IEEE vol No.14,No.4, July,2006.

[10]. Keith Dana Martin, "Sound-Source Recognition: A Theory and Computational Model". Ph.D. thesis, MIT ,1999.

[11]. Fang Zheng, Guoliang Zhang And Zhanjiang Song, "Comparison of Different Implementations of MFCC" , J. Computer Science & Technology, 16(6): 582-589, Sept. 2001.

[12]. Sirko Molau, Michael Pitz, Ralf Schl¨uter, and Hermann Ney, "Computing Mel-Frequency Cepstral Coefficients On The Power Spectrum".