International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 7 5001 - 5006

Classification of Musical Instruments sounds by Using MFCC and Timbral Audio Descriptors

Priyanka S. Jadhav
M.E. (Computer Engineering)
G. H. Raisoni College of Engg. & Mgmt.
Wagholi, Pune, India
E-mail: priyanka10jadhav@gmail.com

***Abstract*—** Identification of the musical instrument from a music piece is becoming area of interest for researchers in recent years. The system for identification of musical instrument from monophonic audio recording is basically performs three tasks: i) Pre-processing of inputted music signal; ii) Feature extraction from the music signal; iii) Classification. There are many methods to extract the audio features from an audio recording like Mel-frequency Cepstral Coefficients (MFCC), Linear Predictive Codes (LPC), Linear Predictive Cepstral Coefficients (LPCC), Perceptual Linear Predictive Coefficients (PLP), etc. The paper presents an idea to identify musical instruments from monophonic audio recordings by extracting MFCC features and timbre related audio descriptors. Further, three classifiers K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) and Binary Tree Classifier (BT) are used to identify the musical instrument name by using feature vector generated in feature extraction process. The analysis is made by studying results obtained by all possible combinations of feature extraction methods and classifiers. Percentage accuracies for each combination are calculated to find out which combinations can give better musical instrument identification results. The system gives higher percentage accuracies of 90.00%, 77.00% and 75.33% for five, ten and fifteen musical instruments respectively if MFCC is used with K-NN classifier and for Timbral ADs higher percentage accuracies of 88.00%, 84.00% and 73.33% are obtained for five, ten and fifteen musical instruments respectively if BT classifier is used.

Keywords- musical *instrument identification; sound timbre; audio descriptors; feature extraction; classification.*

*****

I. INTRODUCTION

Musical instrument identification is one of the most important aspects in the area of Music Information Retrieval (MIR). The musical instrument identification by machine becomes the area of interest recently as most of the music is available in digital format. The music can be available in various textures like monophonic, polyphonic, homophonic, heterophonic, etc. The monophonic texture includes sound of only one musical instrument. The biphonic texture consists of two different musical instruments sounds played at the same time. In polyphonic texture sounds of multiple musical instruments are include which are independent from each other to some extent. The homophonic texture is the most common texture in western music. It contains multiple musical instruments sounds played at a time which are dependent on each other, so differs from the polyphonic texture. The heterophonic texture contains two or more sounds of musical instruments which are played simultaneously performing variations of the same melody. It is most challenging to identify musical instruments from a music piece involving more than one instrument playing at the same time which is referred as polyphonic audio but the great deal of work still has to be carried out in the monophonic or solo context [1], [2].

The proposed work deals with the identification of musical instrument from a monophonic audio sample where only one instrument is played at a time. Sounds produced by same musical instrument have similar features. This music related features are extracted from sound samples by using different feature extraction methods. There are many methods to extract characteristics or features from audio samples. Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Codes (LPC), Linear Predictive Cepstral Coefficients (LPCC), Perceptual Linear Prediction (PLP) are mostly used techniques for audio feature extraction. In this paper, with traditional MFCC feature extraction method we also focused on extracting timbre related attributes from sound samples. Audio descriptors that are used to extract timbral characteristics from audio files are addressed in [3]. These audio descriptors are discussed later in this paper.

The audio features extracted from sound samples by using same feature extraction method are compared with each other on the basis of some algorithm called as classifier, to find similar sounds. Various classifiers like Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), K-Nearest Neighbor (KNN), Bayesian classifiers, Artificial Neural Networks (ANN) etc. can be used for classification process. In proposed system we are working with three different classifiers namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Binary Tree Classifier (BT) to identify musical instrument. The purpose of proposed work is to achieve two objectives: (a) to identify musical instrument by extracting mfcc and timbral attributes from sound sample and (b) to analyze which feature extraction method and classifier can gives better identification results. Further in this paper we have discussed the concept of audio descriptors and sound timbre, timbre related audio descriptors, our proposed system, results and conclusion.

II. LITERATURE REVIEW

The huge research exists in area of Music Information retrieval (MIR) is mainly concentrated on speaker recognition musical instrument identification and singer identification [4], [5]. Machine recognition of musical instrument is quite recent area for research. The majority of work deals with identifying musical instrument from monophonic sound sources consisting of only one instrument playing at a time. Much work was initially dedicated to propose relevant features for musical instrument identification in [6], [7], [8], which basically includes temporal features, spectral features, and cepstral features as well as their variations. In further work the effect of combining features for musical instrument identification was studied as in [6], [9]. Various feature extraction methods, audio descriptors and classifiers useful for musical instrument identification are studied by some researchers in [10], [11], [12]. In [13] different classification techniques along with their accurecy rates for instrument identification are studied.

K-Nearest Neighbor (KNN) classifier is most commonly used by many researchers in their work for instrument identification in solo context [14], [15], [16]. Discriminant analysis is used in [17] and in [18] decision trees are used as for classification purpose. Artificial Neural networks (ANN/NN) are also used in many studies like [15], [19]. Gaussian mixture models (GMMs) and hidden Markov models (HMMs) were also considered by some researchers as in [1], [15], [20], [21], [22]. The support vector machines (SVMs) [1], [13], [23] are also found successful for instrument identification.

With this, the other work needed to be considered is the research related to timbre recognition. Timbre can be considered as a quality of sound which enables us to distinguish between two sounds. Various definitions and terms related to timbre are discussed in [10]. However till now very small work is done on the identification of musical instruments by using timbral attributes of sound. Many researchers work on musical instrument identification by recognizing sound timbre and also present their work on audio descriptors that are useful for extracting timbre related characteristics from audio file as in [3], [8], [10], [24], [25], [26]. Audio descriptors can be considered as the characteristics or attributes of sound. Audio descriptors describe the unique information of an audio segment [4]. Two sound samples of same musical instrument have similar features. The set of audio descriptors extracted from an audio single can uniquely define it and make it differentiable from other audio signals. A music sound can be described by four factors: pitch, loudness, duration, and timbre [10]. The pitch, loudness and duration are all one dimension entities while timbre is multidimensional in nature.

Till now, no one is able to define the term “timber” accurately. The pitch can be measured in Hzs, loudness can be measured in dB, duration can be measured in seconds but the timbre has no unit of measurement. Timbre is a quality of sound by which we are able to distinguish between two sounds, which are of same pitch, loudness and duration.

Many researchers gave their comments on timbre. Number of definitions and comments about timbre which are given by researchers are discussed in [10]. We have summarized some definitions here.

In, [27] Fletcher defines timbre as: “Timbre depends principally upon the overtone structure; but large changes in the intensity and the frequency also produce changes in the timbre”. Licklider comments in [28] that, “It can hardly be possible to say more about timbre than that it is a 'multidimensional' dimension”. In [29] Helmholtz use term ‘tone quality’ as alternative to the timbre and define it as, “the amplitude of the vibration determines the force or loudness, and the period of vibration the pitch. Quality of tone can therefore depend upon neither of these. The only possible hypothesis is that the quality of tone should depend upon the manner in which the motion is performed within the period of each single vibration”. An American Standards Association (ASA) defines timbre as, “timbre is that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar” [10].

III. TIMBRE RELATED AUDIO DESCRIPTORS

Some audio descriptors that are considered for extracting timbre related characteristics from audios in [3] are: Attack time, Attack slope, Zero Crossing Rate (ZCR), Roll off, Brightness, MFCC, Roughness and Irregularity, which are described below:

A. Attack time

An attack phase is described by using attack time. The temporal duration of an audio signal is estimated by the attack time. Attack time is the way a sound is initialized [26].

B. Attack slope

The attack slope gives the average slope of the attack time. The values are expressed in same scale as the original signal but they are normalized by time in seconds. It specifies method for slope estimation.

C. Zero Crossing Rate (ZCR)

The noisiness of sound is represented by the Zero Crossing Rate (ZCR). It is measured by counting number of times the audio signal changes its sign. If the sound signal has less number of sign changes then the value of Zero Crossing Rate (ZCR) is smaller for that signal. However for the noisy sound, Zero Crossing Rate (ZCR) will be high.

D. Roll off

Roll off is a way to measure amount of high frequency in the sound signal. It is calculated by finding the frequency in such a way that certain fraction of total energy is always contained below that frequency. This ratio is fixed to 0.85 by default.

E. Brightness

The brightness is similar to the roll off. The cut-off frequency is fixed first and the brightness is calculated by measuring amount of energy above that cut-off frequency. The value of brightness is always in between 0 to 1.

F. MFCC

Mel Frequency Cepstral coefficients (MFCC) describe the spectral shape of an audio input. It is a multiprocessing system. First, the frequency bands are logarithmically positioned. This is called as Mel scale. A method that has energy compaction capability called as Discrete Cosine Transform (DCT) is used, that considers only the real numbers. By default first 13 components are taken.

G. Roughness

Roughness is an estimation of sensory dissension. It represents a rapid sequence of important events occurring in the audio sample. Roughness of a sound depends on the shapes of the events and the frequency of occurrence of those events. Roughness values are higher when short duration events occur for a fixed pulse frequency, while it is smaller when the pulse frequency is higher.

ISSN: 2321-8169
5001 - 5006

H. Irregularity

Irregularity is the degree of variation of the sequential peaks of the spectrum. It is sum of square of the difference between amplitudes of neighboring partials. Optionally, there is another approach to find the irregularity. It is calculated as the sum of amplitude minus mean of previous, same and next amplitude.

From these we are going to use only six audio descriptors for feature extraction in our proposed system. The audio signals, which are inputted to system are of fixed duration and contain continuous amplitude throughout the signal. Hence, there is not much significance in considering the attack time or attack slope for feature extraction in our research.

IV. Proposed System

The proposed system deals with three steps as given below:

i) Preprocessing of musical instrument sound sample,

ii) Extraction of audio features from the sound sample by using (a) traditional MFCC method and (b) non-traditional timbral feature extractors;

iii) Classification using K-Nearest Neighbor (K-NN), Support Vector Machine (SVM) and Binary Tree (BT) classifiers.

In first step, the musical instrument sound sample which is in solo context is taken as an input to a system. The database is maintained which contains all these normalized sound samples per musical instruments. In next step, our work deals with both traditional MFCC feature extraction method as well as non-traditional timbral feature extractors. The timbre related audio descriptors are already explained in previous section of this paper. The set of extracted audio descriptors is then used to generate a feature vector.

Three classifiers K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) and Binary Tree (BT) are used to identify the musical instrument. Among this the K-Nearest Neighbors is most popular statistical classifier used by many researchers for classification of musical instruments. Further in third step, classification is done. The block diagram of proposed system is shown in fig.1.

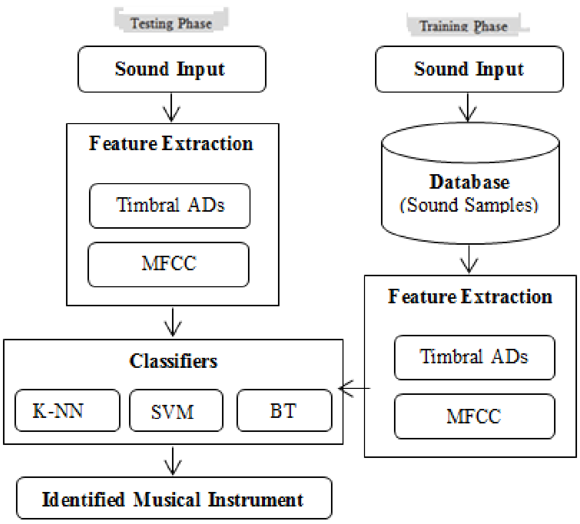

Fig.1: Block diagram of proposed system

The system works with two phases, (i) training phase and (ii) testing phase. In training phase, known sound samples are given as input to system. All features are extracted from these samples by using one feature extraction methods and placed in a matrix or vector format called as features vector. One classifier is trained by using given features vector for further classification process. KNN classifier does not require training. In testing phase an unknown sound sample is given as an input to system and related features of music signal are extracted by using same feature extraction method which is used in training phase. These features are then compared with the reference features obtained in training phase and the new signal is then classified by using same classifier.

The purpose of our proposed work is to achieve two objectives: (a) to identify musical instrument by extracting timbral attributes from sound sample and (b) to analyze which feature extraction method and classifier can gives better identification results. To achieve second objective, percentage accuracies are calculated by making all possible combinations of feature extraction methods and classifiers.

V. Database

Database is maintained with sound samples of fifteen musical instruments. All audio samples are the wave files with same duration and properties. Twenty-five such sound samples are collected per each of the fifteen musical instruments. From these fifteen samples each are used for training and ten samples each are used for testing purpose. The properties of collected sound samples are given below:

1. Audio File Type: Wave sound (.wav)
2. Texture: Monophonic
3. Frequency: 11025 Hzs
4. Bit rate: 16 bits/sec
5. Duration: 3 seconds

TABLE I: Database

Sr. No.	Musical Instrument Name	Sr. No	Musical Instrument Name
1.	BANSURI	9.	PICCOLO
2.	BENJO	10.	PIYANO
3.	SITAR	11.	SANTOOOR
4.	CLARIONET	12.	SARANGI
5.	GUITAR	13.	SAROD
6.	HARMONIUM	14.	SAXOPHONE
7.	ISRAJ	15.	SHEHANAI
8.	NADSWARAM		

VI. EXPERIMENTS AND RESULTS

Experiments are made by making all possible combinations of feature extraction methods and classifiers. In this manner total six experiments are done for different number of musical instruments.

TABLE II: EXPERIMENTS PERFORMED WITH FIVE MUSICAL INSTRUMENTS

Experiment No.	Feature Extraction Method	Classifier	Percentage Accuracy (%)
1.	MFCC	K-NN	90.00%
2.	MFCC	SVM	82.00%
3.	MFCC	BT	92.00%
4.	Timbral ADs	K-NN	72.00%
5.	Timbral ADs	SVM	82.00%
6.	Timbral ADs	BT	88.00%

The percentage accuracy for each experiment shown in TABLE II is calculated for first five musical instruments in TABLE I. The combinations of MFCC with BT classifier and Timbral ADs with BT classifier are giving maximum percentage of accuracies of 92.00% and 88.00% respectively.

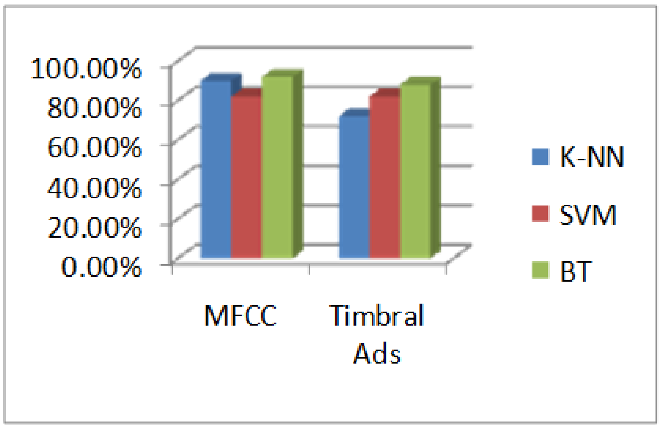

Fig.2: Percentage accuracies obtained for five musical instruments.

TABLE III: EXPERIMENTS PERFORMED WITH TEN MUSICAL INSTRUMENTS

Experiment No.	Feature Extraction Method	Classifier	Percentage Accuracy (%)
1.	MFCC	K-NN	77.00%
2.	MFCC	SVM	64.00%
3.	MFCC	BT	71.00%
4.	Timbral ADs	K-NN	54.00%
5.	Timbral ADs	SVM	50.00%
6.	Timbral ADs	BT	84.00%

The percentage accuracy for each experiment shown in TABLE III is calculated for first ten musical instruments in TABLE I. The combinations of MFCC with K-NN classifier and Timbral ADs with BT classifier are giving maximum percentage of accuracies of 77.00% and 84.00% respectively.

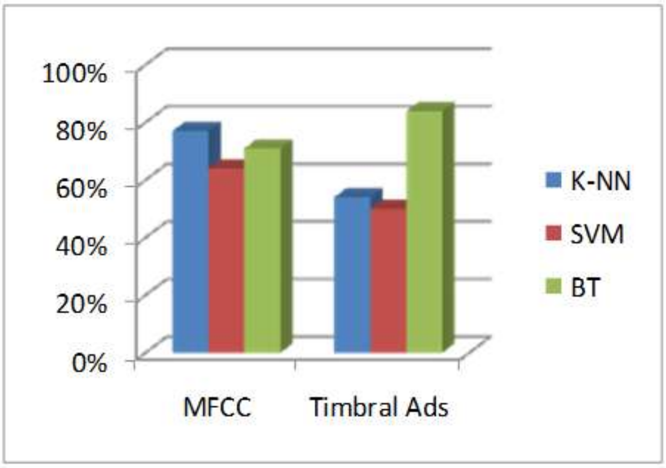

Fig.3: Percentage accuracies obtained for ten musical instruments.

TABLE IV: EXPERIMENTS PERFORMED WITH FIFTEEN MUSICAL INSTRUMENTS

Experiment No.	Feature Extraction Method	Classifier	Percentage Accuracy (%)
1.	MFCC	K-NN	75.33%
2.	MFCC	SVM	60.33%
3.	MFCC	BT	66.66%
4.	Timbral ADs	K-NN	50.66%
5.	Timbral ADs	SVM	46.66%
6.	Timbral ADs	BT	73.33%

The percentage accuracy for each experiment shown in TABLE IV is calculated for all fifteen musical instruments in TABLE I. The combinations of MFCC with K-NN classifier and Timbral ADs with BT classifier are giving maximum percentage of accuracies of 75.33% and 73.33% respectively.

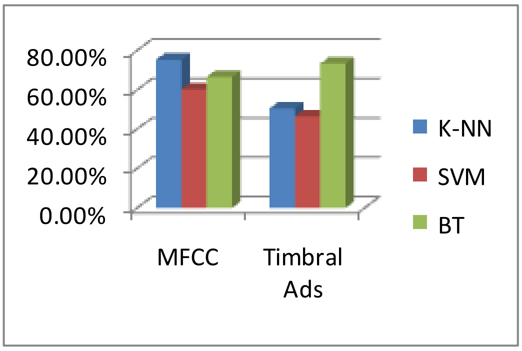

Fig.4: Percentage accuracies obtained for fifteen musical instruments.

The graph for combinations of feature extraction methods and classifiers giving highest percentage accuracies for classification of five, ten and fifteen musical instruments sounds is shown in fig 5.

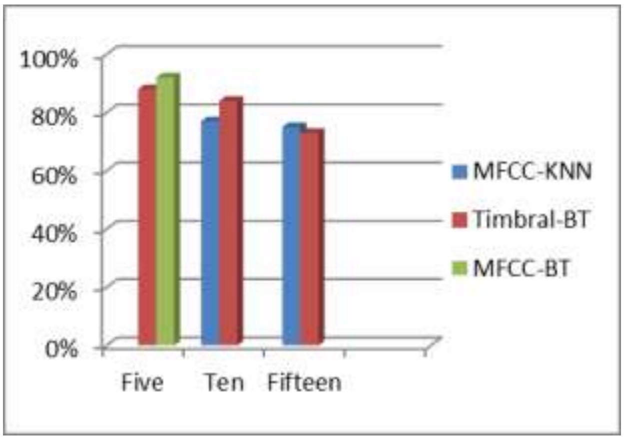

Fig.5: Highest percentage accuracies obtained for five, ten and fifteen musical instruments.

VII. CONCLUSION

The proposed system deals with recognition of musical instruments from monophonic audios. The music related features are extracted from audio samples by using timbral feature extractors as well as traditional MFCC feature extraction method. Three different classifiers namely K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) and Binary Tree (BT) are used to identify musical instrument from a sound sample. The system gives maximum percentage accuracies of 92.00% and 88.00% for combinations of MFCC with BT classifier and Timbral ADs with BT classifier respectively; for five musical instruments. MFCC with K-NN classifier and Timbral ADs with BT classifier give maximum percentage accuracies of 77.00% and 84.00% respectively; for ten musical instruments. For fifteen musical instruments; MFCC with K-NN classifier and Timbral ADs with BT classifier give maximum percentage accuracies of 75.33% and 73.33% respectively. By studying all results one can conclude that the proposed system gives higher accuracy for MFCC if K-NN classifier is used and for Timbral ADs if BT classifier is used.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R.B.G.) thanks . . .” Instead, try “R.B.G. thanks”. Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

REFERENCES

[1] S. Essid, . G. Richard and . B. David, "Musical Instrument Recognition by Pairwise Classification Strategies," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 14, no. 4, JULY 2006.

[2] D. Giannoulis and . A. Klapuri, "Musical Instrument Recognition in Polyphonic Audio Using Missing Feature Approach," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 21, no. 9, SEPTEMBER 2013.

[3] O. Lartillot, MIRtoobox 1.5 Users Manual, August 2013.

[4] S. H. Deshmukh and . 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," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, no. 2, February 2014.

[5] S. H. Deshmukh and 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," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 3, no. 2, February 2014.

[6] S. Essid, G. Richard and B. David, "Efficient musical instrument recognition on solo performance music using basic features," in *AES 25th International Conference*, London, U.K., 2004.

[7] A. Eronen, "AUTOMATIC MUSICAL INSTRUMENT RECOGNITION," TAMPERE UNIVERSITY OF TECHNOLOGY.

[8] X. Zhang and Z. . W. Ras, "Analysis of Sound Features for Music Timbre Recognition".

[9] A. Eronen, "COMPARISON OF FEATURES FOR MUSICAL INSTRUMENT RECOGNITION," Tampere, Finland.

[10] T. H. Park, "Towards Automatic Musical Instrument Timbre Recognition," Princeton University, Deprtment of Music.

[11] D. M. Chandwadkar and M. . S. Sutaone, "Selecting Proper Features and Classifiers for Accurate Identification of Musical Instruments," *International Journal of Machine Learning and Computing,* vol. 3, no. 2, April 2013.

[12] P. Herrera, X. Amatriain, E. Batlle and X. Serra, "Towards instrument segmentation for music content description: acritical review of instrument classification techniques".

[13] P. Herrera, G. Peeters and . S. Dubnov, "Automatic Classification of Musical Instrument Sounds," *Journal of New Music Research,* vol. 32, 2003.

[14] A. Glowacz, W. Glowacz and A. Glowacz, "Sound Recognition of Musical Instruments with Application of FFT and K_NN Classifier with Cosine Distance".

[15] S. K. Banchhor and A. Khan, "Musical Instrument Recognition using Spectrogram and Autocorrelation," *International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307,* vol. 2, no. 1, March 2012.

[16] P. Shinde, . V. Javeri and . O. Kulkarni, "Musical Instrument Classification using Fractional Fourier Transform and KNN Classifier," *International Journal of Science, Engineering and Technology Research (IJSETR),* vol. 3, no. 5, May 2014.

[17] G. Agostini, M. Longari and E. Pollastri, "Musical instrument timbres classification with spectral features," in *Proc. International Workshop Multimedia Signal Processing*, Cannes, France, Oct. 2001.

[18] K. Jensen and J. Amspang, "Binary Decision Tree Classification of Musical Sound," in *ICMC Proceedings*, 1999.

[19] G. Mazarakis, . P. Tzevelekos and . G. Kouroupetr, "Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks".

[20] M. Eichner, M. Wolff and . R. Hoffman, "Instrument classification using Hidden Markov Models".

[21] A. Eronen, "MUSICAL INSTRUMENT RECOGNITION USING ICA-BASED TRANSFORM OF FEATURES AND DISCRIMINATIVELY TRAINED HMMS".

ISSN: 2321-8169
5001 - 5006

[22] J. C. Brown, "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features," *J. Acoust. Soc.,* vol. 105., Mar. 1999.

[23] C. N. Copeland and . S. Mehrotra, "Musical Instrument Modeling and Classification".

[24] R. Moore, "COMPUTER RECOGNITION OF MUSICAL INSTRUMENTS: AN EXAMINATION OF WITHIN CLASS CLASSIFICATION," SCHOOL OF COMPUTER SCIENCE AND MATHEMATICS, VICTORIA UNIVERSITY, June 2007.

[25] T. Zhang, "Instrument Classification in polyphonic music based on timbre analysis".

[26] S. H. Deshmukh and S. G. Bhirud, "Analysis of Audio Descriptor Contribution in Singer Identification Process," *International Journal of Emerging Technology and Advanced Engineering,* vol. 4, no. 2, February 2014.

[27] H. Fletcher, "Loudness, Pitch and Timber of Musical Tones and their Relations to the Intensity, the Frequency and the Overtone Structure," *JASA,* vol. 6, no. 2.

[28] J. C. R. Licklider, Basic Correlates of the Auditory Stimulus, New York: Wiley.

[29] H. L. Helmholtz, On the Sensation of Tone as a Physiological Basis for the Theory of Music, New York: Dover Publications.