2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India

Musical Instrument Identification using MFCC

Monica S. Nagawade,
Department of Electronics &
Telecommunication, GECA, Aurangabad
monicanagawade2410@gmail.com

Varsha R. Ratnaparkhe,
Department of Electronic &
Telecommunication, GECA, Aurangabad,
patwadkar.varsha@gmail.com

Abstract- Music signal processing is one of the active research area now a days. Identifying the musical instrument form the solo recordings is one of the applications under signal processing. Different types of algorithms have been proposed till now to identify the instrument. For this objective first of all we have to extract the features from sound samples and then we can use them for instrument identification. In this paper we are presenting one of efficient method to extract the features from sound sample that is Mel Frequency Cepsrtal Coefficient (MFCC). These MFCCs are closely related to behavior of human auditory system. While calculating these MFCCs we actually deal with spectral envelop. And hence it gives very distinctive features of the sound samples. So using these features for identification of musical instrument we can increase the accuracy of identification. In our work we have used five instruments and 30 samples of each instrument. In the training database total 90 samples (60% of total samples) and for testing database 60 (40%)samples are used. In the classification phase we have used K- Nearest Neighbor (K-NN) classifier. Proposed system gives the accuracy of 91.66% for Cello, Piano and Trumpet and accuracy of 83.33% for Flute and Violin.

Keywords - Musical instrument identification, signal processing, mel frequency cepstral coefficients, K-Nearest Neighbors, feature extraction, Timber.

I. INTRODUCTION

Application of signal processing for the music and speech analysis has been increasing widely. There are several types of applications under music analysis such as extraction of melody from music sound, recognition and separation of sound sources in polyphonic audio, recognition of musical instrument in the isolation, Automatic music transcription, Beat tracking, Musical information retrieval and many more. Any sound sample from the musical instrument can have number of attributes associated with it. If we successfully extract these features from that sound sample then it will be possible for us to use those features to identify that particular instrument. Music has many different features such as pitch (fundamental period), harmony (simultaneous presentation of pitches), timber etc.

In the proposed work we are dealing with one of the application of signal processing i.e. Identification of musical instrument from the solo recordings. One of the features of musical sample i.e. Mel frequency cepstral Coefficients are considered for instrument identification. MFCCs are basically used to describe the timber of sound. These MFCCs are widely used for both speaker recognition and musical instrument recognition. Timber is attribute with which two sounds having same loudness and pitch can be distinguished. Olivier Lartillot and Petri Toiviainen et.al [9] [10] showed that timber is also having other attributes associated with it such as attack time, attack slope, zero crossing rate, RMS energy, centroid, flatness, roll-off, irregularity, brightness and roughness etc. Using MIRToolbox as given we can extract the musical features for classification. The extraction of features from sound sample can be useful in other applications such as musical information retrieval. In this extracted features can be given as query to retrieve the data of interest.

In this paper we propose the methods to extract the features from sound samples and based on these features instrument is identified. The basic system block diagram is as shown in figure 1. Sound samples for the proposed system have been taken from Electronic Music Studio, University of Iowa [11]. Database available is created by playing instruments with more than one technique, including arco, pizzicato, vibrato, and non-vibrato. For proposed work we have used samples having characteristic acro and non-vibrato. Acro means insturments are played with bow. And non-vibrato means the fundamental frequency (Pitch) is constant. Samples are recorded with frequency 44.1 kHz at resolution of 16 bits. The instruments used are given in table I below.Properties of sound samples:

1. Frequency: 44.1 KHz
2. Bit Rate: 16 bits/sec
3. Audio File type: Wave file(.wav)
4. Texture: Monophonic
5. Duration: 1-2 seconds

TABLE I: Database

Sr. No.	Name of Musical Instrument
1	Piano
2	Flute
3	Violin
4	Trumpet
5	Cello

978-1-5090-3704-9/17/$31.00 © 2017 IEEE

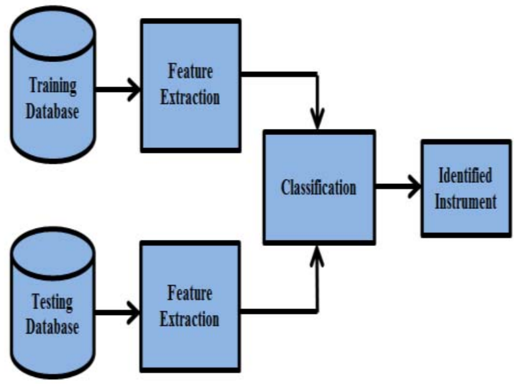

Figure 1: Basic system block diagram

II. RELATED WORK

Different types of methods by many researchers have been proposed for the Identification of Musical Instrument. It is also evident from the literature review that this field is still having research scope. Some of the feature extraction and classification methods are discussed below.

Meinard Müller, Daniel P. W. Ellis, Anssi Klapuriand Gaël Richardet.al [1] presents paper on signal processing techniques for music analysis. This paper summarizes various research fields and the proposed woks in the field of music signal processing. The use of MFCCs for the musical instrument identification is also overviewed. So this paper gives the idea about how choose the specific area for the research.

D. G. Bhalke , C. B. Rama Rao and D. S. Bormane et.al [2]presents MFCCs based on Fractional Fourier transform for classification of musical instruments. Discrete Fractional Fourier transform is used to calculate features. Temporal related features such as zero crossing rates, attack time, decay time, energy are also computed. Zero crossing rates is calculated by,

$$\mathrm{ZCR} = \frac{1}{T}\sum_{0}^{T-1} |\mathrm{sgn}[x(y)] - \mathrm{sgn}[x(y-1)]| \quad (1)$$

x(y) - signal at y^{th} sample
x(y-1) - signal at (y-1) sample
T - samples in one frame

The energy of the of the sound sample is given by,

$$\mathrm{Energy} = \sum_{n=0}^{T-1} (|m(n)|)^2 \qquad (2)$$

m(n) - signal at n^{th} sample
T- samples in one frame

Priyanka S. Jadhav et.al [3] proposed the use of MFCC and Timbral related audio Descriptors for the musical instrument identification. For feature classification k- nearest neighbor, support vector machine and binary tree are used. Identification accuracies for different combinations of features extraction and classification method are compared.

Dattatraya Kuralkar and Saurabh Deshmukh et.al [4] presents a method for musical instrument identification with the use of audio descriptors. In the proposed work sound samples from five different flute instruments is collected. For the feature extraction MIR toolbox used. Then system is able to identify the sound of particular flute instrument.

Slim Essid, Gael Richard, and Bertrand David et.al [5] proposed use of MFCC features of sound samples. Delta MFCC features are used which are obtained by taking time derivative of MFCCs. Spectral features such as spectral centroid, spectral width etc. are also computed. For the classification SVM algorithm is used. In SVM one versus one mapping is used to create trained data.

M. E. Ozbek, N. Ozkurt, and F. A. Savaci et.al [6] proposed a technique of instrument identification using wavelet ridges. For this first 3 level wavelet decomposition is performed as shown in figure 2. This wavelet decomposition gives one approximate coefficient and three detailed coefficients.

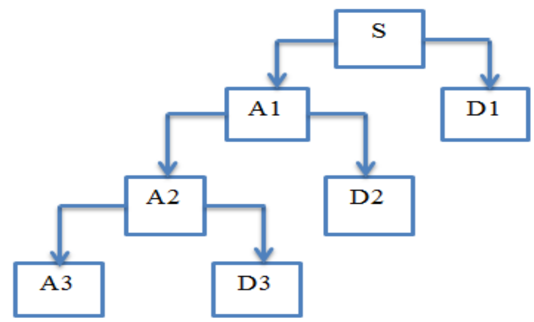

Figure 2: 3 level wavelet decomposition

S= A3+D3+D2+D1

S= signal frame,
A= approximate coefficient, D=detailed coefficient.

Farbod and Karthikeyan et.al [7] proposed musical instrument classification using wavelet dependent time scale features. In this first continuous wavelet transform of the signal frame is taken and then features related to temporal variation and bandwidth are considered for feature extraction.

Sumit Kumar Banchhor and Arif Khan et.al [8] proposed a method of musical instrument identification using short term energy and ZCR. For this analysis first signal is divided into frames and for each frame energy and ZCR is computed. ZCR is nothing but count of how many times signal changes the sign.

III. PROPOSED WORK

Musical instrument identification has important application in musical information retrieval system. Our proposed system is capable of identifying which musical instrument has played based on its sound samples. The proposed work can identify the musical instrument from the solo sound samples.

The proposed work has been divided into two parts: 1) Feature extraction and 2) Classification. In feature extraction MFCCs of the sound samples are obtained and in classification phase it is identified that which instrument was played. The flow of the proposed work is as described below.

A) Feature Extraction:

In this phase Mel Frequency Cepstral Coefficients are extracted from the sound sample. We have chosen these features because they closely resembles the human auditory system. Human auditory system does not give equal response to entire frequency ranges. And hence using these features we can acquire highest recognition accuracy. The flow of feature extraction is as shown in figure 3.

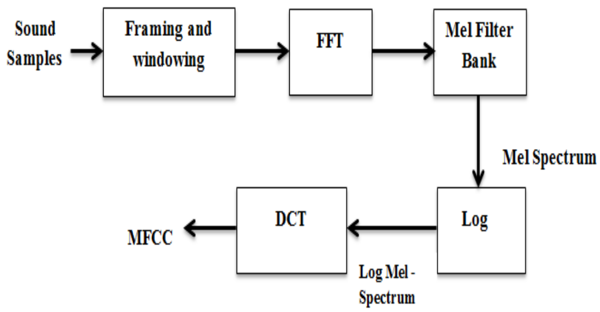

Figure 3: Block diagram of MFCC

a. Framing and Windowing: In this step sound sample is segmented into the frames. Individual frame is consisting of 441 samples. After this hamming windowing is done to smoothly attenuate both ends of the signal towards zero. Hamming window reduces the amplitudes of both sides, so FFT is more focused on the middle part in time domain. Therefore 50% overlapping of windows is done to cover everything equally. Hamming windowing is chosen because it has higher selectivity. The plot of Hamming window for 441 samples in time domain is shown in figure 4.

b. FFT of each frame: Fast Fourier Transform is calculated for converting individual frame of 441 samples from time domain to frequency domain.

c. Mapping of the spectrum to Mel scale :In this Triangular Filter bank is created on Mel frequency scaling. Filter bank is shown in figure 5. And then applied to spectrum. In this stage 22 channels are created to obtain 22 MFCCs. Conversion from frequency in Hetrz to mel is done using

$$\mathrm{fr} = 2595 \times \log_{10}(1 + (\mathrm{f} \div 700))(1)$$

f - Real frequency

fr - Perceived frequency

d .Logarithmic conversion :Human response to signal level is logarithmic. Therefore after applying the Mel filter bank to the individual frames of signal its logarithmic conversion is performed. This conversion shrinks dynamic range.

e. Discrete Cosine Transform: After applying DC The log Mel spectrum is get converted into time domain. The result is nothing but Mel frequency Cepstrum Coefficients (MFCC). These MFCCs are then stored as feature vectors. The MFCC plot for piano note A4 is shown in figure 6.

B) Classification (K-NN):

In this phase Euclidean Distance Metric is used to compute the distance of test data feature vector with that of training data. It is calculated by:

$$d(a,b) = \sqrt[2]{\sum_{i=1}^{n}(bi - ai\ )}\ \hat{}\,2(2)$$

a-Testing data feature vector.

b-Training data feature vector.

After this trained data samples with minimum distance to test data are used for classification. In the K-NN first fivenearest neighbors are considered and among them the one with highest score gives the identification result.

We have created the GUI for proposed system. With Which individuals can easily understand and access the system. GUI is presented in figure 7.

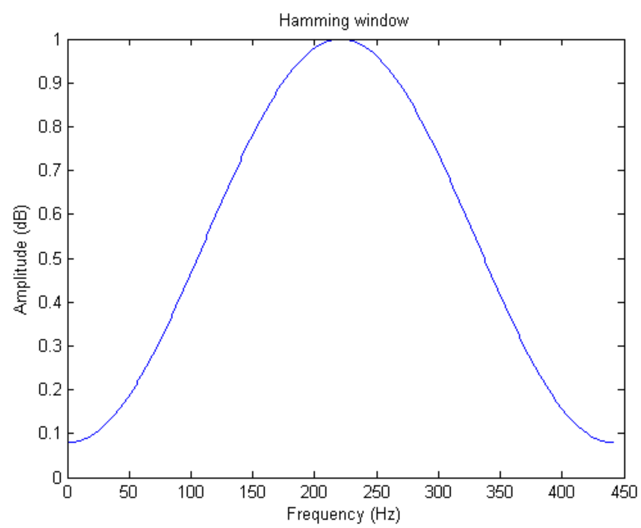

Figure 4: Plot of hamming window

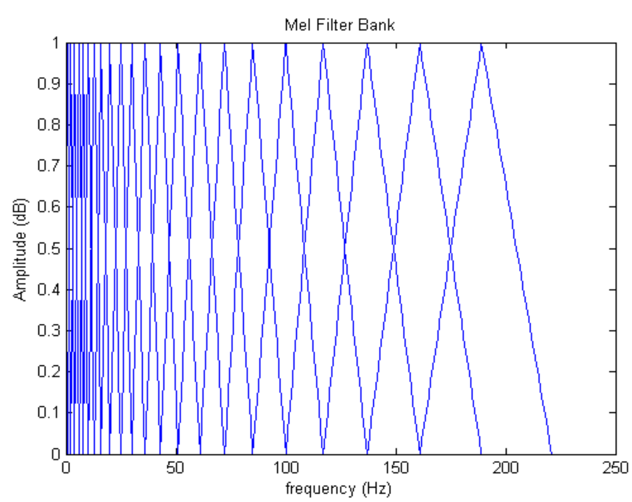

Figure 5: Plot of Mel filter bank

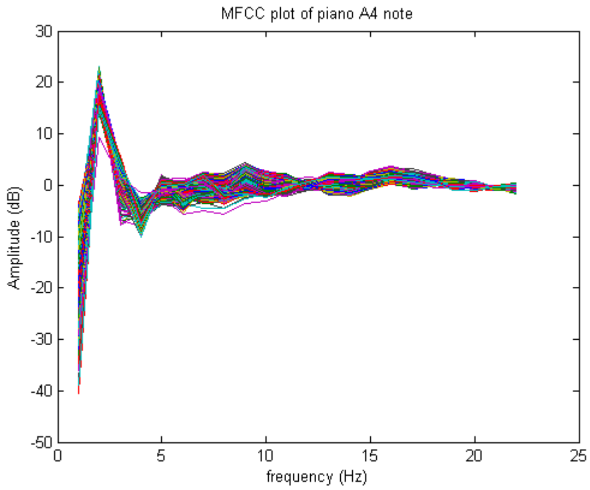

Figure 6: Plot of MFCC coefficients for piano A4 note

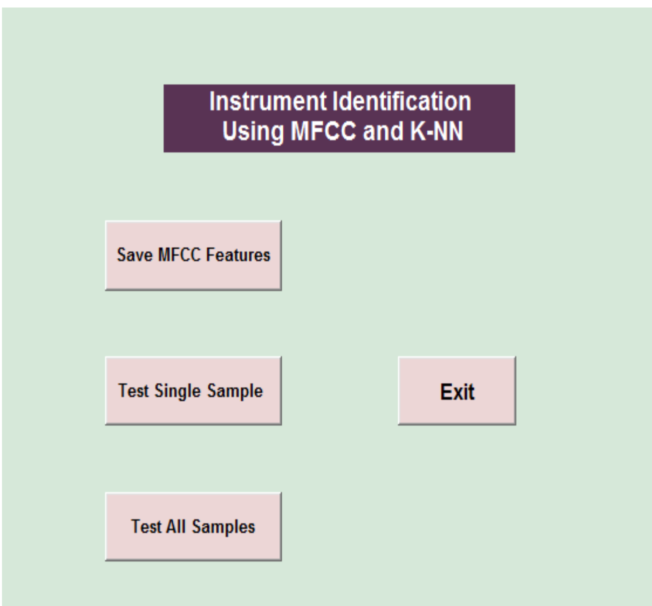

Figure 7: GUI of the proposed system

IV. RESULT

In the proposed system in the feature extraction phase we have calculated the MFCC features for all the sound samples in the training database. In this vectors are get created which contains frame wise MFCC features for all the sound samples. In the proposed system the total training samples are 90 so in the MFCC feature data we get total 90 feature vectors.

While testing any sound sample for recognizing musical instrument we have first calculated its MFCC feature vector and then using Euclidean distance metric we have calculated difference of that test data with all the training feature vectors.

While performing actual experiments on all the test samples which 60 in number, most of the sound samples get correctly classified and some of them get wrongly classified. These results are shown in table II. Depending on accurately classified sound samples we calculated the percentage accuracy of identification for each instrument. This is shown in table III.

Table II: Result of instrument identification

Sr. No.	Name of Musical Instrument	Number of samples used for testing	Wrongly classified samples
1	Cello	12	1
2	Flute	12	2
3	Piano	12	1
4	Trumpet	12	1
5	Violin	12	2

Table III: Accuracy of instrument identification

Sr. No.	Name of Musical Instrument	Accuracy of Identification (%)
1	Cello	91.66
2	Flute	83.33
3	Piano	91.66
4	Trumpet	91.66
5	Violin	83.33

V. LIMITATIONS

1. The proposed system will not be able to identify the musical instrument in the polyphonic sound recordings.
2. As we go on increasing number of instruments for classifications recognition accuracy will start decreasing.

CONCLUSION

In the proposed system MFCC approach is used for feature extraction from sound samples and K-NN classification is used for classifying instrument sound. For Cello, Piano and Trumpet recognition accuracy is 91.66% and for Flute and Violin it is 83.33%.

From the results of proposed system we conclude that identification accuracy is good with the MFCC. But if use more number of feature extraction method and different classifiers then system accuracy can be improved with more number of musical instruments used.

REFERENCE

[1] Meinard Müller, Member, IEEE, Daniel P. W. Ellis, Senior Member, IEEE, Anssi Klapuri, Member, IEEE, and Gaël Richard, Senior Member, IEEE. "Signal processing for music analysis". IEEE Journal of selected topics in signal processing, VOL. 5, NO. 6, OCTOBER 2011

[2] D. G. Bhalke , C. B. Rama Rao and D. S. Bormane, "Automatic musical instrument classification using fractional Fourier transform based- MFCC features and counter propagation neural network" , J. Intell. Inf. Syst. (2016) 46:425–446.

[3] Priyanka S. Jadhav, " Classification of Musical Instruments sounds by Using MFCC and Timbral Audio Descriptors"IJRITCC, ISSN: 2321-8169, Volume: 3 Issue: 7, 5001 – 5006

[4] Dattatraya Kuralkar, Saurabh Deshmukh, "Study of Audio Descriptors for Specific Musical Instrument Identification in North Indian Classical Music", IJSR ISSN (Online): 2319-7064.

[5] Slim Essid, Ga¨el Richard, and Bertrand David "Efficient musical instrument recognition on solo performance music using basic features", AES 25th international conference, london, united kingdom, 2004 june 17–19

[6] M. Erdal Ozbek , Nalan Ozkurt and F. Acar Savaci, " Wavelet ridges for musical instrument classification",J Intell Inf Syst (2012) 38:241–256, DOI 10.1007/s10844-011-0152-9

[7] Farbod Foomany and Karthikeyan Umapathy, "Classification of music instruments using wavelet-based time-scale features", 2013 IEEE ICMEW.

[8] Sumit Kumar Banchhor and Arif Khan , "Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy", IJAIS – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 1– No.3, February 2012 – www.ijais.org.

[9] Olivier Lartillot and Petri Toiviainen, "A matlab toolbox for musical feature extraction from audio", Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, 2007.

[10] Olivier Lartillot and Petri Toiviainen, "mir in matlab (ii): a toolbox for musical feature extraction from audio".

[11] Electronic Music Studio, University of Iowahttp://theremin.music.uiowa.edu/MIS.html.