International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 5, May 2014

Musical Instrument Classification using Fractional Fourier Transform and KNN Classifier

Pravin Shinde , Vikram Javeri , Omkar Kulkarni

Abstract— **The classification of musical instrument, where the idea is to build a system that will listen to musical note and recognize which instrument is being played. In recognition of musical instrument the different musical instrument played based on isolated monophonic notes. Different feature component will be extracted and will be given to the classification system. Especially we will use the feature component based on Fractional Fourier Transform. The k-nearest neighbors will be used as a classifier for recognition. System will recognize which instrument is being played and it belong to which family will be displayed on GUI. This system proved that FRFT gives more accuracy of classification.**

Index Terms- **Fractional Fourier Transform, k-nearest neighbors, Musical Instrument, recognition**

I. INTRODUCTION

Automatic musical instrument recognition is a crucial subtask in solving these difficult problems, and may also provide useful information in other sound source recognition areas, such as speaker recognition. Musical signal analysis do not have much commercial interest as, except, speaker and speech recognition. This is because the topics around speech processing are more readily commercially applicable, although both areas are considered as being complicated. Through constructing computer systems that "listen", we may also gain some new insights into human perception. This thesis describes the construction and evaluation of a musical instrument recognition system that is able to recognize single tones played by an instrument.

A central concept in our study is the quality of sound, i.e. what something sounds like. Musical sound is said to have four perceptual attributes: *pitch, loudness, duration* 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 counter parts. For musical sounds, pitch is well defined and is almost equal to the fundamental frequency. The physical counterpart of loudness is intensity. Intensity is directly proportional to square of the amplitude of the acoustic pressure. The third dimension of musical signal is, perceived duration, and corresponds quite closely to the physical duration with tones that are not very short. Timbre is the least understood among the four attributes. From past time timbre is defined by exclusion: the quality of a sound by which a listener can tell that two sounds of the same loudness and pitch are dissimilar. We are fortunate in the sense that many researchers have explored the underlying acoustic properties that cause different sound quality, or timbre sensations. Based on this information about music signal, and adding some knowledge about the physical properties of which produce sound instruments, we can try to construct algorithms that measure this information from digitally stored acoustic signals.

Commonly, classification is performed with statistical pattern recognition techniques. In musical synthesis, the model parameters are often analyzed from an acoustic signal. There might be potential in combining these two fields, using physical model synthesis parameters for musical instrument recognition and bringing new methods for feature extraction from musical instrument recognition to physical modeling.

Musical Instrument Classification

Musical Instruments are classified in five different families depend on their shape, method of playing instrument , sound they produce.

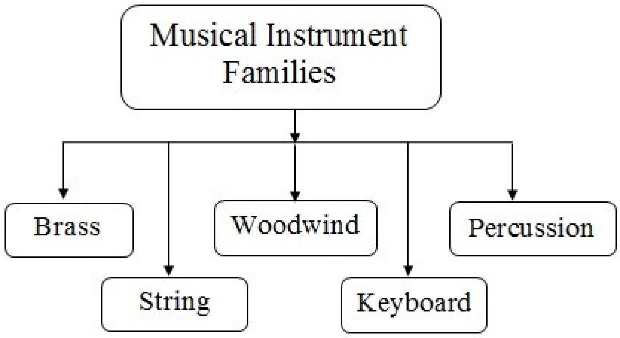


Figure 1. Musical Instrumnet families

1.Brass Instruments

The brass instrument made from a cup-shaped mouthpiece, a slightly tapered mouth pipe, cylindrical tubing, valves, and hyperbolic bell like structure. Player introduces puffs of via vibrating lips stretched over the mouthpiece. The length of the air column is change by tubing switched in with help of valves, either piston or rotary. A common arrangement is such that the first valve lowers the intonation by two semitones, the

Manuscript received April 2014.

This work is supported by Rajarshi Shahu College of Engineering, Pune. The paper was submitted for review on.

Pravin Eknath Shinde Electronics & Telecommunication, Pune University, Pune University, Rajarshi Shahu College of Engg, Pune, India.

Vikram Hemant Javeri Electronics & Telecommunication, Pune University, Pune University, Rajarshi Shahu College of Engg, Pune, India.

Omkar Milind Kulkarni Electronics & Telecommunication, Pune University, Pune University, Rajarshi Shahu College of Engg, Pune, India.

ISSN: 2278 – 7798

second by one semitones, and the third by three semitones.

Examples: French Horn, Trumpet, Tuba etc.

2. String Instrument

Some instrument strings instruments are set into vibration by plucking and initiated and maintained by using bowing. The frequency of vibration is basically established by the tension, mass per unit length, length of the string. A string vibrates at the lowest (fundamental) frequency and same time at higher frequencies which tend toward integer multiples of the fundamental frequency. The radiation of sound from a stringed instrument is enhanced by a resonator consisting closed air cavity. Some part of energy of the vibrating string is transmitted via the bridge to the walls of the cavity.

Examples: Cello, Bass, Violin, Guitar etc.

3. Woodwind Instrument

Woodwind instruments are distinguished by the fact that the effective length of the vibrating air column is shortened by opening lateral side holes in succession. Two different means and distinct way of generation of sound are employed. This method of excitation leaves the tube acoustically open in the sense that the contained air vibrates much as it does in a simple tube with both ends open to the atmosphere. For both the single as well as double-reed instruments, the vibration of reed because sound waves reflected back from the distant end of the air column and allow the puffs of air to enter when the sound pressure within the instrument is large. Examples: Bass ,Oboe, Saxophone etc.

4. Keyboard Instruments

The instrument are played as reeds, pipes, strings vibrating bars, in these instruments are selected by use of keys in a keyboard.
Examples: Harpsichord, Piano, Pipe Organ etc.

5. Percussion Instruments

The sound is initiated by a blow. Two types of sound producers are present, a membrane under tension, which is associated with a cavity which influence the frequency of vibration, plate vibrating transversely, whose frequency is little affected by any resonator that may be attached.
Examples: Steel Drum, Tymapni, Xylophone, etc.

II. System Structure

The system is design for the identification and classification of instruments. In order to classify musical instrument into particular class we need to find out significant information about the signal. This is referred as feature extraction process. The parameter extracted from music signal depends on type of instrument and it's playing style. This significant information will be used in classification process to identify the correct instrument.
The system is divided in two parts: Training and Testing.

1.Training Phase
In this phase set of known signal is used an input. The feature of known signal will be extracted using FRFT and this features placed in a matrix or vector format as a Reference Model which contain standard database for classification

B. Testing Phase
In this phase an unknown test signal will be given as an input and using FRFT feature of the signal will be extracted. This feature will be compared with the reference features. By using classifier we are able to identify which feature matching amongst all feature. We are in position to identify instrument and family.

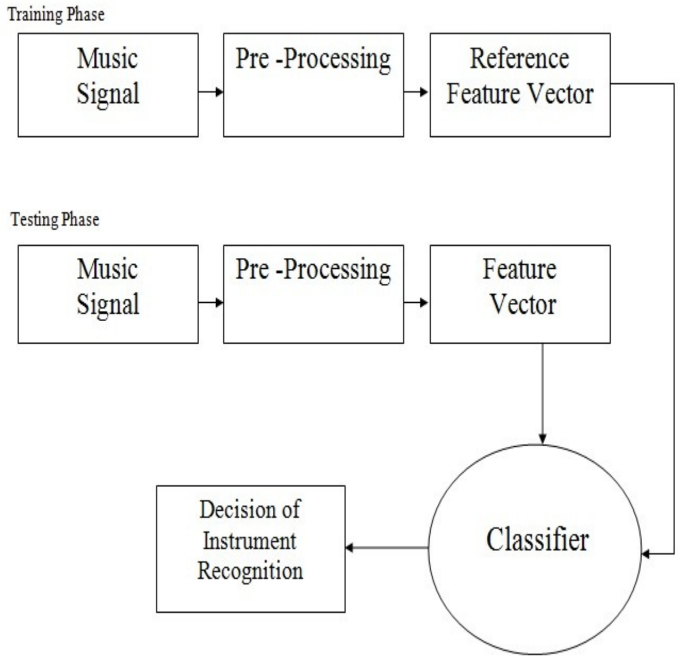


Figure 2. Proposed block diagram

Block diagram consist of -
1. Pre-processing
In this stage noise present in signal, silence part from input signal is removed. Silence present in signal is removed with different methods like Zero Cross detection Rate (ZCR), short time energy distribution based method. After this Framing, windowing is done to remove end effects.

2. Feature extraction
Musical signals have different features. There are different feature extractions techniques are available to find exact feature of signal. We will use FRFT (Fractional Fourier Transform) to extract the features of signal. Features of signal will extract in this step.

3. Reference model
In reference model ,features are stored in matrix format for each instrument. Based on features vector unknown input signal is classified and recognized.

4. Classifier
Classifier has input from reference model and extracted feature of unknown signal. Classifier is used to classify

ISSN: 2278 – 7798

test signal on the basis of reference vector. We used KNN classifier for our system.

The flow of system can be represented by the following flowchart:

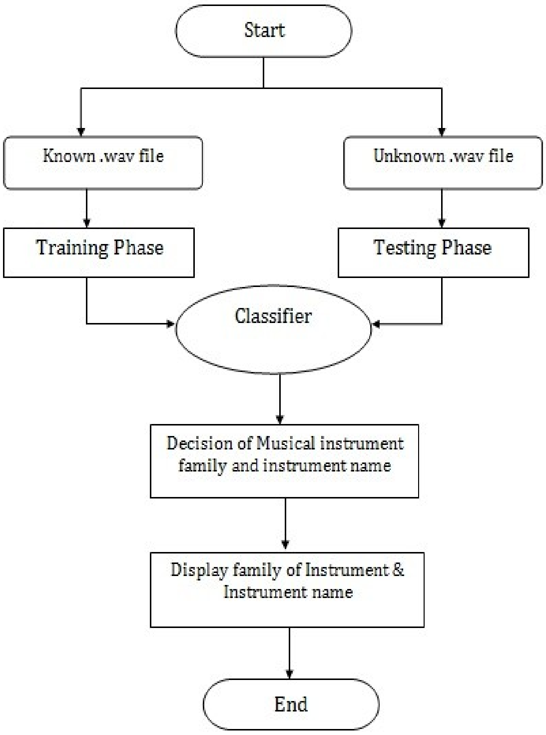


Figure 3.Flowchart of system

III. METHODOLOGY

A. Pre Prosessing

The pre-processing stage involves various amplification and filtering stages to be employed on the given musical signal. Various pre-processing techniques involves:-
1. Removal of silence.
2. Framing.

In silence removal amplitude of music signal is compare with threshold value. Music signal is break into frames. Each frame of frame duration 0.020 seconds which is called frame by frame analysis. In each frame we identify the non silence part by comparing amplitude with 0.04 which is threshold value. New signal is created which does not contain any silence part.
The Algorithm for silence removal part :
Step 1) Break the signal into the frames of 0.020 second frame duration.
Step 2) Identify the non silence part by comparing frame by frame amplitude with threshold=0.04
Step 3) Store non silence part of signal in to new signal

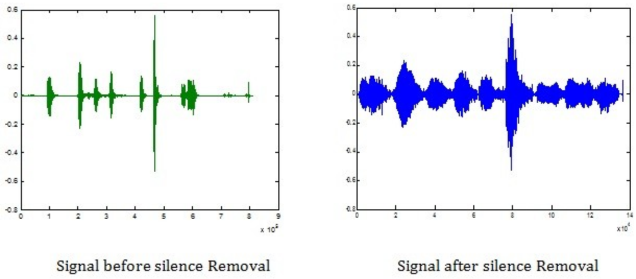


Figure 4. Silence Removal

B. Feature Extraction

Fractional Fourier transform (FRFT) is used for feature extraction. The fractional Fourier transform is a time–frequency distribution and an extension of the classical Fourier transform. FRFT can be think as Fourier transform to the *n*-th power, where *n* need not be an integer. It is function which will transform function to any intermediate domain between time and frequency [15]. It is used to filter noise but condition is that it should not overlap with desired signal in the time domain. The FRFT is a generalization of the ordinary Fourier transform with an order parameter α and is identical to the ordinary Fourier transform when this order a is equal to π/2[13] . The FRFT is a linear operator that corresponds to the rotation of the signal when an angle not a multiple of π/2. FRFT is the representation of the signal along the axis *u* making an angle a with the time axis.
The FRFT is defined with the help of the transformation kernel Kα as,

$$K_\alpha(t,u) = \begin{cases} \delta(t-u) & \text{if } \alpha \text{ is a multiple of } 2\pi \\ \delta(t+u) & \text{if } \alpha+\pi \text{ is a multiple of } 2\pi \\ \sqrt{\dfrac{1-\mathrm{j}\cot\alpha}{2\pi}}\, e^{\mathrm{j}((u^2+t^2)/2)\cot\alpha - \mathrm{j}\,ut\,\mathrm{cosec}(\alpha)} & \text{if } \alpha \text{ is not a multiple of } \pi \end{cases}.$$

The FRFT is defined using this kernel as (FRFT of order α of *x(t)* denoted by *Fα(u)*

$$F_\alpha(u) = \int_{-\infty}^{\infty} x(t) K_\alpha(t,u)\,dt$$

$$= \begin{cases} \sqrt{\dfrac{1-j\cot\alpha}{2\pi}}\, e^{j\frac{u^2}{2}\cot\alpha} \displaystyle\int_{-\infty}^{\infty} x(t) e^{j\frac{t^2}{2}\cot\alpha} e^{jut\csc\alpha}\,dt & \text{, if a is not a multiple of } \pi \\ x(t) & \text{, if a is a multiple of } 2\pi \\ x(-t) & \text{, if } \alpha+\pi \text{ is a multiple of } 2\pi \end{cases}$$

Fractional Fourier transform can be used in time frequency analysis, to filter noise, but with the condition that it does not overlap with the desired signal in the time frequency domain. After Taking FRFT of Signal following features were computed.

ISSN: 2278 – 7798

1.Mel Frequecy Cepstral Coefficient(MFCC)

The MFCC feature can be extracted in six major steps performed in the following order:

1. Framing
2. Windowing
3. Discrete Fourier Transforming
4. Mel-Frequency Warping
5. Log Compression and Discrete Cosine Transforming

1. Framing

The first step is framing. The signal is split up into frames typically with the length of 10 to30 milliseconds. The frame length is important because of tradeoff between timeand frequency resolution. The frames overlap each other typically by 25% to 70% of their own length.

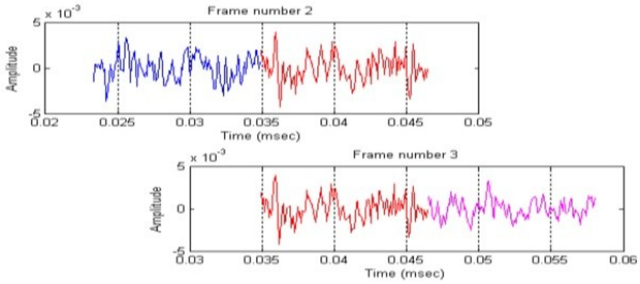


Figure 5. Framming of signal

2. Windowing

After the signal is split up into frames each frame is multiplied by a window function. For good window function main lobe should be narrow and side lobe should be low. A smooth tapering is present at the edges is desired to minimize discontinuities. Hamming window is used to remove end effects. Hamming window is defined as

$$W[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right),\ 0 \le n \le N-1$$

$$=0, \quad \text{otherwise}$$

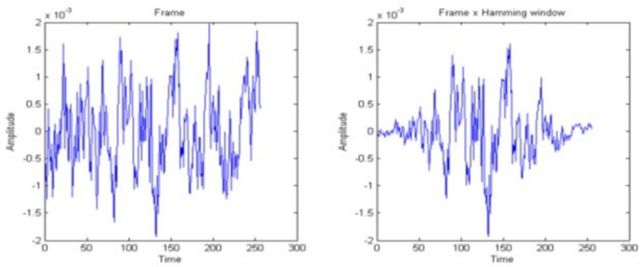


Figure 6. Signal before and after hamming window applied

3. Discrete Fourier Transform

Third step is to take discrete cosine transform of each frame.

$$Xn = \frac{1}{N}\sum_{k=0}^{N-1} Xk \cdot e^{i2\pi kn/N}, (n \in Z)$$

4. Mel-Frequency Warping

The mel scale is based on pitch perception and uses triangular-shaped filter. The scale is roughly linear below 1000 Hz and non-linear (logarithmic) after 1000 Hz .

Mel frequency=2595*log(1+llinear frequency/700)

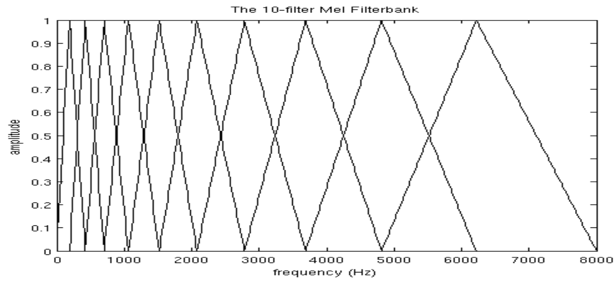


Figure 7. Filter bank of 24 triangular-shaped filter

5. Log Compression and Discrete Cosine Transforming

Last step is compress the spectral amplitude by taking log and discrete cosine transform is computed and this give us cepstral coefficient and by spectral smoothing MFCC coefficient computed.

2.Features

After computing MFCC of signal different features are computed they are as follow:

a)Skewness

It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value may be positive or negative, or even undefined. The skewness of a random variable *X* is nothing but third standardized moment. Skewness is denoted as $\gamma 1$

$$\gamma 1 = E\left[(X - \mu/\sigma)^3\right]$$

b) Varience

The average of the squared differences from the mean. variance measures how far a set of values is spread out. Variance is non-negative. A small variance indicates that the value is very close to the mean (expected value).

$$Var(X) = E\left[(X - \mu)^2\right]$$

c) Standard Deviation

Standard deviation is square root of varience . It indicate the dispersion of value from its mean value.

$$\sigma = \sqrt{Var(X)}$$

d)Spectral Flux

Spectral flux is the rate of change of the power spectrum. It measure how quickly the power spectrum changes from frame to frame. It calculate by comparing the power spectrum of one frame with power spectrum of the previous frame.

$$\text{Spectral flux} = \sum_{k=2}^{K} \left| M(fk) - M(fk-1) \right|^2$$

e)Spectral Centroid

It indicate the location of the centre of gravity of the magnitude spectrum. It gives the impression of 'brightness' of sound. It is evaluated as the weighted mean of the spectral frequencies.

ISSN: 2278 – 7798

$$\text{Spectral Centroid}=\frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

f) Spectral Roll-Off

Spectral roll-off is defined as the frequency bin M below which 85% of the magnitude distribution is concentrated. It is measure of spectral shape.

$$\sum_{n=0}^{M} f(n) = 0.85 * \sum_{n=0}^{N} f(n)$$

g)Kurtosis

It is measure of the peakedness of the probability distribution of the variable. Peakedness is nothing but width of peak.

C. Classifier

KNN is non parametric lazy learning algorithm. Non parametric means it does not make any assumption on the underlying data distribution. Its outstanding characteristic is that it does not require a training stage in the strict sense. The training samples are rather used directly by the classifier during the classification stage. The key idea behind this classifier is that, if we are given a test pattern (unknown feature vector), ***x***, we first detect its *k*-nearest neighbors in the training set and count how many of those belong to each class. In the end, the feature vector is assigned to the class which has accumulated the highest number of neighbors. Therefore, for the *k*-NN algorithm to operate, the following ingredients are required:

1. A dataset of labeled samples, i.e. a training set of feature vectors and respective class labels.

2. An integer $k \geq 1$.

3. A distance (dissimilarity) measure.

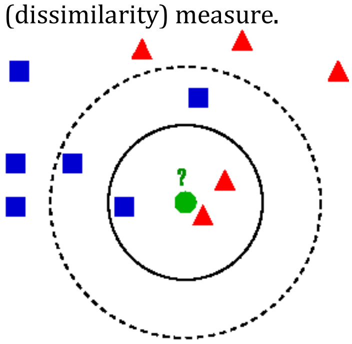


Illustration of KNN algorithm

In the first step ,the algorithm computes the distance, *d(**x**, **vi**)*, between ***x*** and each feature vector, ***vi***, $i = 1,\ldots,M$, of the training set, where *M* is the total number of training samples. The most common choice of distance measure is the *Euclidean distance*, which is calculated as:

$$d(x,vi) = \sqrt{\sum_{j=1}^{D} \left(x(j) - vi(j)\right)^2}$$

where *D* is the dimensionality of the feature vector. After *d(**x**, **vi**)* has been computed for each ***vi*** , the resulting distance values are sorted in ascending order. As a result, the *k* first values correspond to the *k* closest neighbors of the unknown feature vector. Now, let *ki* be the number of training vectors among the *k* neighbors of ***x*** that belong to the *i*th class, $i = 1,\ldots,Nc$ .The unknown vector is then classified to the class which corresponds to the maximum *ki* .Further more,the posterior probability *P(ωi|**x**)* can be estimated as:

$$P(\omega i|\mathbf{x}) = ki/k \qquad i = 1,2\ldots,N_c$$

Therefore, from a Bayesian perspective, the algorithm assigns the unknown sample to the class that corresponds to the maximum estimated posterior probability. Based on these values instruments are classified.

IV. Conclusion

After doing experiments and calculation we conclude some vital observations.

1. Fractional Fourier transform improves accuracy of classification.
2. FRFT based MFCC give best result among all feature than only MFCC.
3. System become more efficient if we add spectral feature to it. (Spectral flux, Spectral roll off, Spectral centroid)
4. If we find accuracy of classification with consideration of respective family of an instrument, all feature provide more improved result than those are with instrument wise classification.
5. If we try to find best feature of an individual instrument then if become very complicated process.

Acknowledgment

We are thankful to the Department of Electronics and Telecommunication, Rajarshi Shahu College of Engineering for all the support it has provided us. We express our deepest sense of gratitude towards our Honorable Head of department and project guide Prof. D.G.Bhalke for giving us permission to use the college resources and his constant encouragement for this work.

REFERENCES

Papers:

[1]George Tzantaki, Perry Cook,"*Music Classification of audio signal*" IEEE Transaction on speech and audio processing vol.No.5 July 2002.

[2] I. Kaminskyj and C. pruysers , *"Wavelet analysis in musical instrument and sound classification "*Proceeding Of IEEE 2009.

[3] Inge Gavat and Mithai Datcu , *"Wavelet analysis for audio signal with music classification "* Proceeding Of IEEE 2009.

[4] Khalid Youssef , *"Music note recognition system based on time delay neural network "* Fourth International Conference On Natural Computation of IEEE 2008.

ISSN: 2278 – 7798

[5] Ketsuro Etahara "*Musical instrument identification based on F0 dependent multi vibrate normal distribution*" Proceeding of IEEE 2003.

[6] Jonghyu Lee and antti Erronen, "*Music instrument recognition using ICA based transforms of features and discriminatively trained HMM's* " Proceeding of IEEE 2003.

[7] Cyril Joder & Slim Essid "*Temporal integration for audio classification with application to musical classification*" IEEE Transaction on Speech and audio processing Vol.No1,Jan 2009.

[8] Tao Li and Qi Li , "*Comparative study on content based music genre classification* " Proceeding of IEEE 2003.

[9]Tao Li and Qi Li , "*Music Feature ectraction using wavelet coefficient histogram* " United State Patent August 2006.

[10]Chang Hsing Lee , Jau-Ling Shih, Kun-Ming Yu, and Hwai-San Lin "*Modulation spectral analysis of traditional information of cepstral & spectral features for music genre classification*" 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

[11]Sumeet Kumar & Arif Khan "*Musical instrument recognition using Zero crossing rate and short time energy*". International Journal of Applied Information system (IJAIS)-S,New York,USA Volume 1-No.3,Feb 2012.

[12] Yibin Zhung "Classification of recorded Musical instrument sound based on Neural Networks"Proceedings of the 2007 IEEE symposium on Computational Intelligence in Image and Signal Processing.

[13] V. Ashok Narayanan, K.M.M.Prabhu "*The fractional Fourier transform: theory, implementation and error analysis*" Microprocessors and Microsystems 27 (2003) 511–521.

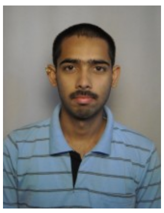
[14] Ricardo A. Garcia, "*MUSICAL INSTRUMENT FAMILY CLASSIFICATION*", Massachusetts Institute of Technology.

[15] L. B. Almeida, "*The fractional Fourier transform and time-frequency representations,*"IEEE Trans. Signal Processing, vol. 42, no. 11, pp. 3084-3091, Nov. 1994.

Books:
[1] MATLAB R2011b, Mathswork.
[2]Dr. Shaila D. Apte , "Speech and Audio Processing" , Wiley India Edition

Pravin E. Shinde is currently pursuing the undergraduate course Bachelors of Engineering in Electronics and telecommunication at Rajarshi Shahu College of Engineering, Pune

Vikram H. Javeri is currently pursuing the undergraduate course Bachelors of Engineering in Electronics and telecommunication at Rajarshi Shahu College of Engineering, Pune



Omkar M. Kulkarni is currently pursuing the undergraduate course Bachelors of Engineering in Electronics and telecommunication at Rajarshi Shahu College of Engineering, Pune

ISSN: 2278 – 7798