International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.1 (2014), pp.275-282
http://dx.doi.org/10.14257/ijsip.2014.7.1.26

# Hybridization of Fractional Fourier Transform and Acoustic Features for Musical Instrument Recognition

D. G. Bhalke[1], C. B. Rama Rao[2] and D. S. Bormane[3]

*[1,2]National Institute of Technology, Warangal, India,*
*[3]JSPM's RSCOE, Pune, India*
*[1]bhalkedg2000@yahoo.co.in, [2]cbrr@nitw.ac.in, [3]bdattatraya@yahoo.com*

***Abstract***

*This paper presents musical instrument recognition for isolated music sound signals using hybridization of fractional fourier transform (FRFT) based features with timbrel (acoustic) features using feed forward neural network. The FRFT based features which is named as fractional MFCC are computed by replacing conventional discrete fourier transform in mel frequency cepstral coefficient (MFCC) with discrete FRFT. Hybrid features are obtained by effectively combining Fractional MFCC with timbrel features such as temporal, spectral and cepstral features. Feed forward neural network with back propagation algorithm has been used to test the performance of system and results were compared in terms of recognition accuracy and number of features. Proposed feature out performs over individual and other traditional features proposed in the literature. The experimentation is performed on isolated musical sounds of 19 musical instruments covering four different instrument families. The system is tested on benchmarked McGill University musical sound database.*

***Keywords:*** *Musical instrument recognition, Mel Frequency Cepstral Coefficient (MFCC), Fractional Fourier transform (FRFT)*

## 1. Introduction

The main objective of this research work is to identify the type of musical instrument and its family from the musical sound using hybrid features using feed forward neural network. Musical instrument recognition has attracted the attention of various researchers because of its many commercial applications like Musical instrument transcription, content-based music retrieval, music genre classification, duet analysis, Musical information retrieval, audio and video retrieval, playlist generation, acoustic environment classification, video scene analysis *etc.*, [1-3].

So far many attempts were made for musical instrument recognition and classification [1]-[5]. Most of them were based on finding effectives features with number of parameters like No. of instruments, type of musical sound, number of features *etc*. The state of the work is briefly described here. The statistical pattern-recognition technique for classification of 15 musical instrument tones with 31 features based on log-lag correlogram was discussed in Martin and Kin [2]. A study on pitch independent musical instrument recognition for 30 musical instruments with 43 features based on spectral, cepstral and temporal properties of sounds was described by Eronen and Klapuri [14]. Tao Li and Qi Li [13] proposed features based on wavelet coefficients at various frequency sub bands of Daubechies wavelet for music genre classification and emotional content of the music. Eronen [1] performed study on musical instrument recognition for comparison of features. Large set of features including MFCC, delta MFCC, Linear prediction cepstral coefficients, temporal features, spectral

ISSN: 2005-4254 IJSIP

features and modulation features for 16 orchestral instruments were used for experimentation in [1]. Kaminskyj and Czaszejko [4] discussed instrument recognition for isolated monophonic notes using six features: cepstral coefficients, constant *Q* transform frequency spectrum, multidimensional scaling analysis trajectories, RMS amplitude envelope, spectral centroid and vibrato for 19 instruments. Deng *et al.*, [3] has discussed study on feature analysis for recognition of classical instruments using different machine learning techniques to select and evaluate features extracted from a number of different feature schemes. The performance of Instrument recognition was analyzed using selected features with different feature selection and ranking algorithms.

Review of earlier work shows that, developing compact and efficient feature set for Musical instrument recognition has become topic of currents research area in Musical instrument recognition and attracted the attention of various researchers. This paper describes about Feature extraction, effective combination of hybrid feature set and experimentation using different combination of feature set. The paper is organized as follows. Proposed system is described in Section 2 and Feature extraction is described in Section 3. Database details are given in Section 4 and performance evaluation with different combination of feature set is given in Section 5. Conclusion is summarized in Section 6 followed by references in Section 7.

## 2. Proposed System

The proposed system for recognition of musical instrument recognition is shown in Figure 1. The system consists preprocessing, feature extraction, and classification. In preprocessing the silence part of the signal is removed by selecting proper threshold value of Zero crossing (ZC) and Energy of signal. Silence removal helps to reduce the computational complexity of the system. In features extraction various acoustic features like MFCC which is most significant and validated for speech and music processing, proposed fractional MFCC feature, Temporal, spectral features are extracted. Different combinations of these features are formed and tested using feed forward Neural Network as classifier.

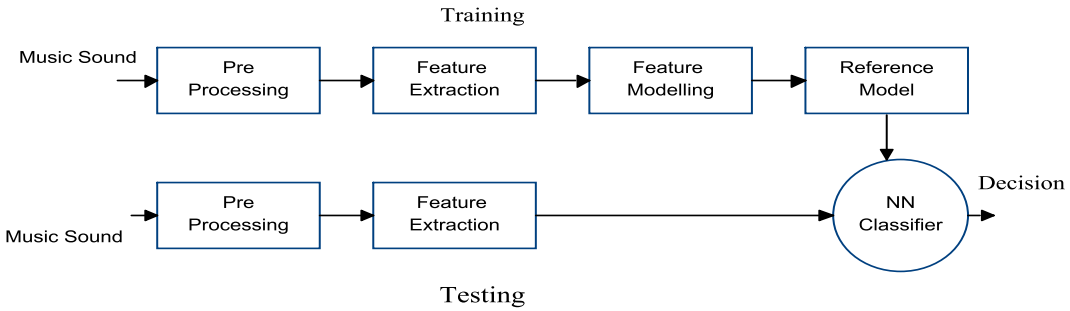


**Figure 1. Proposed Block Diagram**

## 3. Feature Extraction

Extracting most significant features is vital part of any recognition system and is most important part of the system. Features shown in Table 1 have been extracted and described here.

**Table 1. Features used for Experimentation**

| Sr. No. | Name of Features | No. of features | Type of Feature |
|---|---|---|---|
| 01 | MFCC | 12 | Perceptual |
| 02 | Fractional MFCC (Proposed) | 12 | Perceptual |
| 03 | Mean ZCR, STD ZCR, Energy, Log attack time, Attack slope, Decay time, Release time , Sustain time | 8 | Temporal |
| 04 | Mean SC, STD SC, Mean SR, STD SR, Mean SF, STD SF, Mean SS, STD SS | 8 | Spectral |

### 3.1. Mel Frequency Cepstral Coefficient (MFCC):

Mel Frequency Cepstral Coefficients (MFCCs) are cepstral coefficients used for representing audio signal in a way that mimics the physiological properties of the human auditory system. It has proved its significance and extensively used in speech analysis over the past few decades and have more recently received attention in music analysis. MFCC extraction part consists of pre-processing, pre-emphasis, framing, windowing, triangular mel filter bank, log energy and DCT. In pre-processing, silence part of the signal is removed using ZC (Zero crossing) and Energy of the signal by selecting proper threshold value. Then, signal is framed with 20 ms duration with 50% overlapping and windowed with hamming window. Then FFT is computed and passed through bank of 24 triangular mel filter band pass filters. Logarithmic values of these filtered signal is taken and its DCT is computed to de-correlate the MFCC coefficients. Statistical values of these coefficients of all frame are combined and formed 12 coefficients for each note.

### 3.2 Fractional Fourier Transform based MFCC (Fractional MFCC):

Mel Frequency Cepstral Coefficients (MFCCs) is computed in frequency domain. Music sound signal characteristics depend both on time and frequency domain. Also, MFCC are sensitive to noise and its performance degrades in noisy conditions. MFCC has not shown good performance for Music sound signal due to its limitations. Fractional fourier transform, on the other hand transforms the signal in time and frequency plane and captures time and frequency information and removes noise present in the signal. Also certain Music sounds can be better analyzed in time frequency plane. Fractional MFCC features are computed by rotating the signal in time and frequency plane at specific angle using FRFT, so that certain music characteristics will be captured. The Fractional fourier transform (FRFT) is briefly described below.

**3.2.1. Fractional Fourier transform (FRFT):** The FRFT represents the signal in two orthogonal plane of time and frequency axis. It is a linear operator which corresponds to the rotation of the signal between time and frequency plane, where time axis corresponds alpha=0 and frequency axis corresponds to alpha= $\pi/2$ [6],[7]. FRFT is more flexible and suitable for non-stationary signal as compared to fourier transform because of its orthonormal basis of chirp signals and degree of freedom of rotation of time frequency axis, [6-8].

The $\alpha^{th}$ order fractional Fourier transform $F^{\alpha}(u)$ of f(t) is given by equation 1 to 3. Factional Fourier transform is general case of the Fourier transform with similar properties of Fourier transform such as, Linearity, additivity, commutatively, associatively, Time shift, Modulation, Multiplication, differentiation, Parseval's theorem *etc*.

$$F^{\alpha}(u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} e^{j\frac{u^2}{2}\cot\alpha} \int_{-\infty}^{\infty} e^{-j\frac{t^2}{2}\cot\alpha} e^{-jut\csc\alpha} f(t)dt & \text{if } \alpha \neq N\pi,\ N \text{ is integer} \quad (1) \\ f(u) & \text{if } \alpha = 2N\pi,\ N \text{ is integer} \quad (2) \\ f(-u) & \text{if } \alpha = (2N+1)\pi,\ N \text{ is integer} \quad (3) \end{cases}$$

Since $\propto$ varies from 0 to 1, the FRFT of f (t) changes from the time domain ($\propto$ = 0) to the frequency domain ($\propto$ = 1). Different value of $\propto$ , provides additional flexibility and degree of freedom for processing of non-stationary signals [15]. The block schematic of FRFT based MFCC (Fractional MFCC) coefficients are shown in Figure 2. In MFCC the FFT has been substituted by Discrete FRFT and modified MFCC features have been proposed. Significant features are extracted from the signal at specific angle in time and frequency plane. We have extracted FRFT based MFCC features at different values of α. Finally α is set to 0.95 through experimentation due to higher recognition accuracy at this value.

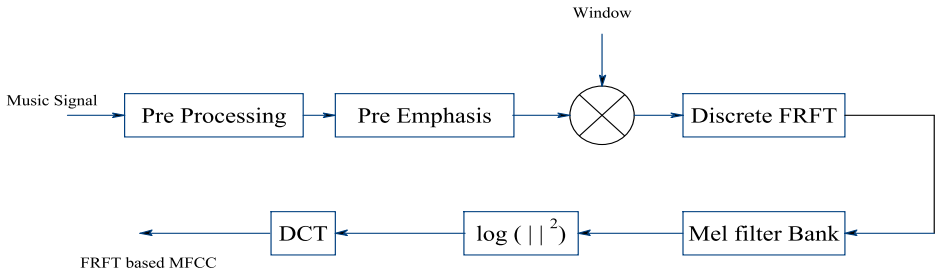


**Figure 2. Block Schematic of Fractional FRFT Features**

**3.3. Timbrel Features:** Timbre, also known as sound quality or tone color of music. It is defined as, when two sounds are heard that match for same pitch, same loudness, and same duration, and a difference can still be heard between the two sounds, that difference is called timbre [10,16]. There are two physical correlates of timbre: spectrum envelope and amplitude envelope**.** Following feature based on spectral and temporal envelope are extracted.

### 3.4. Spectral Features:

*Spectral Centroid (SC):* This is the amplitude-weighted average, or centroid, of the frequency spectrum, which can be related to a human perception or brightness of the instrument [11]. The spectral centroid is given by

$$SC = \left(\frac{\sum_{k=1}^{K} P(f_k) f_k}{\sum_{k=1}^{k} P(f_k)}\right) \quad (4)$$

Where $P(f_k)$ is magnitude spectrum of k$^{th}$ sample and $f_k$ is frequency corresponding to each magnitude element

 

*Spectral flux (SF)*

This is a measure of the amount of local spectral change. This is defined as the squared difference between the normalized magnitude spectra of successive frames and given by equation 2.

$$SF = (\sum_{k=2}^{k} \left| P(f_k) - P(f_{k-1}) \right|) \quad (5)$$

*Spectral spread (SS)*

The spectral spread is a measure of variance (or spread) of the spectrum around the mean value μ .and is given by equation 3.

$$SS = \frac{\sqrt{\sum_{K=0}^{N/2} (P(f_k) - SC)^2}}{\sqrt{\sum_{k=0}^{N/2} (P(f_k)^2}} \quad (6)$$

where $P(f_k)$ = magnitude spectrum corresponding to each magnitude element and SC=spectral centroid.

*Spectral skewness(SK)*

The skewness is a measure of the asymmetry of the distribution around the mean value. The skewness is calculated from the 3rd order moment.

$$SK = \frac{\sum (freq - SC)^2 \times Mag}{\sum Mag} \quad (7)$$

where mag= magnitude spectrum, freq =frequency corresponding to each magnitude element and SC=spectral centroid.

The statistical values *i.e.*, mean and standard deviation of SC,SF, SK, SS were computed and used as timbrel features.

### 3.5. Temporal Features:

Temporal features are extracted in time-domain. Following temporal features are used for experimentation.

*Energy:* It is the sum of the amplitudes present in frame and is defined as:

$$Energy = \sum_{0}^{N-1} (x(n))^2 \quad (8)$$

where *x*[n] is the amplitude of the sample.

*Zero-Crossing Rate:*

This is the number of times the signal crosses zero amplitude during the frame, and can be used as a measure of the noisiness of the signal. It is defined as:

$$ZCR = \frac{1}{N} \sum_{0}^{N-1} | \text{sgn}[x(n)] - \text{sgn}[x(n-1)] | \quad (9)$$

where *sign* = 1 for positive arguments and 0 for negative arguments

*Log-Attack Time:*

The log-attack time is the logarithm of time duration between the time the signal starts to the time it reaches its stable part. It can be estimated by taking the logarithm of the time from the start to the end of the attack.

*ADSR envelope:* Every musical sound are characterized by its temporal envelope which are characterized by Attack time, Decay time , Sustain time and release time. The ADSR values are computed and used as feature vector.

### 3.6. Proposed Hybrid Features:

Aim of our proposed technique is to obtain hybrid features set to get good recognition accuracy. Here, different combination of MFCC, Fractional MFCC, temporal and spectral features are effectively combined and tested for instrument recognition using feed forward neural network. The performance of the proposed automatic music instrument recognition system is analysed in terms of recognition accuracy and number of features.

## 4. Database Details

The dataset used for experimentation is from MUMS (McGill University Master Samples), which is set of 3-DVDs created by: Frank Opolko Joel Wapnick [9]. It is library of isolated music sound tones from a wide number of musical instruments, played with different articulation styles, covering entire pitch and recorded with 44.1 KHz sampling frequency as wave file. Experimentation is done on 760 monophonic isolated notes of 19 musical instruments covering string, Brass, Woodwind and percussion families. 70% notes have been used for training the system and 30% notes for testing the system with cross 10-fold validation method. Instruments used for experimentation are listed in Table 2.

**Table 2. Instrument Used**

| Sr. No | Family | Instruments used |
|---|---|---|
| 1 | String | Guitar, Violin, Viola, Cello, Bass, Lute, Piano, harpsichord |
| 2 | Woodwind | Saxophone , Oboe classical , Oboe D, English Horn |
| 3 | Brass | Trumpet , Tuba, Cornet, Trombone , French Horn |
| 4 | Percussion | Steel drum, Tympani |

## 5. Performance Analysis

In this section, the different combination of hybrid features are effectively combined and evaluated in terms of recognition accuracy and number of features using feed forward neural network as classifier. Recognition accuracy and number of features using neural network is shown in Table 3.Waikato environment for knowledge environment (WEKA) tool has been used for neural network classification [11, 12]. Result shows that recognition accuracy has increased form 75% for MFCC to 94.68% for proposed hybrid feature set which is combination of Fractional MFCC and timbrel feature.

 

**Table 3. Recognition Accuracy in % for Different Classifiers**

| Feature Combination | No. of features | Recognition accuracy (%) |
|---|---|---|
| MFCC | 12 | 75 |
| Timbrel | 16 | 73.13 |
| Fractional MFCC | 12 | 91.84 |
| MFCC + Timbrel | 28 | 90.15 |
| **Fractional MFCC + Timbrel** | **28** | **94.68** |

## 6. Conclusion

In this paper, hybrid features based on combination of FRFT and Timbrel features have been proposed for musical instruments recognition using feed forward neural network. In addition to this, Fractional MFCC features which is short time spectral representation of signal in time and frequency plane has been proposed. Proposed features outperforms over MFCC and other traditional features because of additional degree of freedom of rotation of signal in time and frequency plane. From this work it can be concluded that music sound classes can be better represented in fractional fourier domain.

## References

[1] A. Eronen, "Comparison of features for Musical instrument recognition", In proceeding of IEEE workshop Applications of signal processing to audio and acoustic, **(2001)**, pp. 19-22.
[2] K. D. Martin and Kin, "Musical Instrument recognition: A pattern recognition approach", Journal of Acoustical Society of America, vol. 109, **(1998)**, pp. 1068.
[3] J. D. Deng, C. Simmermacher and S. Cranefield, "A study on Feature analysis for Musical Instrument Classification", IEEE Transaction on Systems, Man and Cybernetics, vol. 38, no. 2, **(2008)**, pp. 429-438.
[4] I. Kaminskyj and T. Czaszejko, "Automatic Recognition of Isolated Monophonic Musical Instrument Sounds using k-NNC", Journal of Intelligent Information Systems, vol. 24, no. 2-3, **(2005)**, pp. 199-221.
[5] G. Agostini, M. Longari and E. Pollastri, "Content-Based Classification of Musical Instrument Timbres", IEEE signal processing society, **(2003)**.
[6] V. A. Narayan and K. M. M. Prabhu, "The fractional Fourier transform: theory, implementation and error analysis", Int. Journal of microprocessors and Microsystems, vol. 27, no. 10, pp. 511-521.
[7] H. M. Ozaktas, Z. Zalevsky and M. A. Kutay, "The fractional Fourier transform with applications in optics and signal processing", New York: Wiley, **(2001)**.
[8] V. Namias, "The fractional order Fourier transform and its application to quantum mechanics", IMA journal of Appl Math, vol. 25, no. 3, **(1980)**, pp. 241-265.
[9] "Mc gill University Master Sample: www.music.mcgill.ca/resources/mum/.html/mums.html.
[10] G. Agostini, M. Longari and E. Poolastri, "Musical instrument timbres classification with spectral features", EURASIP J. Appl. Signal Process, doi: 10.1155/ S1110865703210118, no. 1, **(2003)**, pp. 5-14.
[11] P. K. Ajmera and R. S. Holambe, "Fractional Fourier transform based features for speaker recognition using support vector machine", Int. Journal of Computer and electrical engineering, **(2012)**.
[12] H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques", 2nd ed. San Francisco, CA: Morgan, Kaufmann, **(2005)**.
[13] J. R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Mateo, Appendix: Springer, **(1993)**.
[14] A. Eronen and A. Klapuri, "Musical Instrument Recognition using cepstral coefficients and Temporal features", ICASSP, **(2000)**.
[15] R. Essid and D., "Hierarchical Classification of Musical Instruments on Solo Recordings", Proceedings of ICASSP, **(1988)**.
[16] B. Kostek, "Musical instrument classification and duet analysis employing music information retrieval techniques", Proc. IEEE, vol. 92, no. 4, **(2004)**, pp. 712-729.
[17] T. Li, Q. Li and M. Ogihara, "Music feature extraction using Wavelet coefficient histograms", US Patent 7,091,409 B2, **(2006)**.

## Authors

**Bhalke D.G.**, received B.E. degree from Aurangabad University and M.E. degrees from the Shivaji University Kolhapur in 1998 and 2005 respectively. He is currently pursuing towards the Ph.D. degree at the National Institute of Technology Warangal, India. He is also with the Rajarshi Shahu College of Engineering, Pune, India, where he is an Assistant Professor in the Department of Electronics and Telecommunication Engineering. His research interests include Speech processing and Music signal Processing. He is a life member of Indian Society for Technical Education (ISTE) and Institution of Electronics & Telecommunication Engineers (IETE) society.

**Rama Rao C.B.**, received Ph.D. degrees from IIT Kharagpur, India. Presently he is working as Associate Professor at National institute of technology, Warangal, India. His research interests include Signal processing.

**Bormane D.S.**, received the B.E. degree from Aurangabad University, India in 1987, M.E. degree from Shivaji University, Kolhapur in 1997 and Ph.D. degrees from S.R.T.M. University Nanded, India in 2003. Presently he is working as Professor at Rajarshi Shahu College of Engineering, Tathawade, Pune, India. His research interests include Signal and Image processing. He is a Fellow member of the Institution of Electronics and Telecommunication Engineers, India, Life member of Indian Society for Technical Education (ISTE), New Delhi, Life member of Indian Society for Continuing Engineering Education, (ISCEE) Roorkee, Senior Member of IACSIT, Singapore, Member of Computer Society of India (CSI).