

Musical Instrument Classification using Higher Order Spectra and Hierarchical Taxonomies

F I Kazi
E&TC Dept
JSPM's RSCOE, Pune

D G Bhalke
E&TC Dept
JSPM's RSCOE, Pune

ABSTRACT

In this paper classification of musical instruments is done using higher order spectra using different hierarchical taxonomies. Since last three decades higher order statistics techniques have been developed but until not applied to musical instrument classification. Present work uses higher order spectra to detect and quantify non gaussianity and nonlinearity along with MFCC.

General Terms

Pattern Recognition, Hierarchical classification

Keywords

Higher order spectra, Bispectrum, Hierarchical taxonomies.

1. INTRODUCTION

Music data analysis and retrieval has become very popular research field in the era of digital world. Musical Instrument classification is an important task of music data information retrieval. Instrument recognition depends on great deal on hearing transients of a tone, measuring the beginning (attack), and the ending (release) since they have noise like properties influencing their subjective quality. Music signals are generated with nonlinearity dynamics. To improve prediction and coding quality of such signals then more information available signal must be used which cannot be derived from first and second order statistics such as mean and variance. These tools are useful for analyzing signals from linear processes. The implicit assumption in using these tools is the distribution is normal. The present work mainly focuses on nonlinearity and non-gaussianity of musical instrumental signal.

Glenn Eric Hall, Hassan Ezzaidi, Mohammed Bahoura[1] used static feature vector for direct classification and dynamic feature vector at each node of classification tree is used for hierarchical classification with sequential backward selection (SBS) algorithm feature vector is optimized to keep only best discriminating factors for the instrument identification task. The results obtained with hierarchical classification that improves the score almost to 2% compared to the direct classification. RWC Music Database for musical instrument sound is chosen with isolated notes of instruments. Alicja A. Wieczorkowska, Zbigniew W. Ra's, Xin Zhang, Rory Lewis [10] done hierarchical classification using Hornbostel-sachs classification and by fifteen different articulation method (blown, bowed, bowed vibrato, concussive, hammered, lip vibrated, muted, muted vibrato, percussive, picked, pizzicato, rubbed, scraped and shaken). Hornbostel-Sachs method gives better result than articulation method. Slim Essid, Ga'El Richard & Bertrand David[12] selected the feature which has the first rank is selected by Linear Discriminant Analysis (LDA). Analysis and comparison of two hierarchical taxonomies natural instrument families and automatically hierarchical clustering using SVM classifiers are done. Automatic taxonomy performed only slightly better than the natural one. If non-linear SVM

classifiers are used, as complex decision surfaces can be learned then it is possible to generate natural classification. Rui Rui and Changchun Bao [4] used a novel supervised learning algorithm which classifies string instruments with great accuracy as compared to woodwind instruments. In this the orthogonal basis matrix could be obtained without updating the matrix iteratively, which supervised non negative matrix factorization (NMF) algorithm is unable to do.

Saima Anwar Lashari, Rosziati Ibrahim and Norhalina Senan[3] used a soft set classifier to classify traditional Pakistani string, woodwind and percussion family instruments with perception based and MFCC features for different adaptively parameters such as audio length, frame size and starting point of audio file. Woodwind family performed well than string and percussion. S.Gunasekaran, K.Revathy[8] classified 10 Indian musical instruments from string, woodwind & percussion family with solo and duet recordings. Features are extracted from fractal dimension below 1.95. The feature set includes spectral, perceptual and temporal features of the musical signal. KNN classifier and multi-layer perceptron MLP classifier are used. KNN classifier with $k=11$ gives better result when comparing with MLP classifier. Overall recognition rate is 89.7% for solo and 82.8 % for duet. Akram Azarloo, Fardad Farokhi [6] classified sound of combinations of instruments played simultaneously with UTA algorithm for feature selection. A Perceptual feature signal to mask ratios gives efficient result for KNN & MLP classifier. Roisin Loughran, Jacqueline Walker, Michael O'Neill, Marion O'Farrell [7] classified musical instruments with several features. Best classification seen in centroid data across one octave. Increasing the range decreased the accuracy in classification but still gave encouraging results for pursuing classification across natural range of instruments

2. HIGHER ORDER SPECTRA

First and second order statistics such as mean, variance, autocorrelation, power spectrum are popular signal processing tools. These tools are useful for linear and Gaussian Processes, but when data is deviated from gaussianity and linearity these tools found shortcomings for analysis. Nonlinear and non-Gaussian data can be studied using higher order spectra (HOS). HOS is mainly used to extract information due to deviation from gaussianity, to recover phase information and to detect and quantify nonlinearities in time series. The following sections briefly describe HOS terms described by M.A.A.S. Choudhury and et.al. [14]

2.1 Cumulants

For any random variable x , the moment generating function can be defined as the expectation of the transformation e^{tx} , where $t \in R$, i.e.

$$Mx(t) = E[e^{tx}] \quad (1)$$

Moments can be obtained from the coefficients of the Taylor's series expansion of the moment generating function about the origin. In a similar fashion, Cumulants can be

derived from the coefficients of Taylor's series expansion of cummulant generating function $C_{ox}(t)$, which is defined as the logarithmic of the moment generating function:

$$Cx = \triangleq \ln(Mx(t)) \quad (2)$$

Cumulants are another set of statistical measures which can be used instead of moments because of their excellent noise suppressing properties. Moments and Cumulants are very closely related to each other. They hold following relationship:

$$\begin{aligned} c_1 &= m_1 \\ c_2 &= m_2 - m_1^2 \\ c_3 &= m_3 - 3 m_2 m_1 + 2 m_1^3 \\ c_4 &= m_4 - 4 m_3 m_1 - 3 m_2^2 + 12 m_2 m_1^2 - 6 m_1^4 \end{aligned} \quad (3)$$

Where m_i is the i^{th} order moment defined as $m_i = E[x_i]$ and c_i is the i^{th} order cumulant.

2.2 Bispectrum and Bicoherence

Statistical analysis of time domain signal may not provide all information content of signal. So transforming signal data from time domain to frequency domain can expose the periodicities of the signal and can also aid in understanding the signal generating process. Power spectrum is the frequency domain representation of second order moment. The bispectrum is the frequency domain representation of third order cumulant. The bispectrum is defined as

$$B(f_1, f_2) = E[X(f_1) X(f_2) X^*(f_1 + f_2)] \quad (4)$$

Where $X(f)$ is Fourier transformation of the data series $x(t)$. Equation 4 shows that it is complex quantity having both magnitude and phase. The bispectrum can be plotted against two independent frequency variables, f_1 and f_2 in three dimensional plots. Each point in plot represents bispectral content of the signal at the bifrequency (f_1, f_2) . In fact, the bispectrum at point $(B(f_1, f_2), f_1, f_2)$ measures the interaction between frequencies f_1 and f_2 . This interaction can be related to the non-linearity present in the signal generating systems and therein lies the core of its usefulness in the detection and diagnosis of non-linearity. In this work bispectrum is normalized which is called as bicoherence and is defined by following equation:

$$bic^2(f_1, f_2) \triangleq \frac{|E[B(f_1, f_2)]|^2}{E[|X(f_1)X(f_2)|^2]E[|X(f_1 + f_2)|^2]} \quad (5)$$

where 'bic' is bicoherence function. (Kim and Powers, 1979) shown that the variance of the bicoherence estimator satisfies the following expression:

$$var[bic^2(f_1, f_2)] \approx \frac{1}{M} [1 - bic^2(f_1, f_2)] \quad (6)$$

where M is defined as the number of segments used in the estimation. Bicoherence function value always bounded between 0 and 1.

2.3 Test of Linearity based on Bispectrum

Let $x(k)$ be a discrete stationary time series. We call it linear, if it can be represented by

$$x(k) = \sum_n h(n)e(k-n) \quad (7)$$

where, $e(k)$ is a sequence of independent identically distributed random variables with $E[e(k)] = 0$, $\sigma_e^2 = E$

$[e^2(k)]$, and $\mu_3 = E[e^3(k)]$. For this case, the following frequency domain relationship can be obtained: the power spectrum,

$$P(f) = \frac{\sigma_e^2}{2\pi} |H(f)|^2 \quad (8)$$

And the bispectrum

$$B(f_1, f_2) = \frac{\mu_3}{(2\pi)^2} H(f_1)H(f_2)H^*(f_1 + f_2) \quad (9)$$

where, $H(f) = \sum_n h(n)e^{-inf}$. Equation 5 can be rewritten as

$$\begin{aligned} & bic^2(f_1, f_2) \\ & \triangleq \frac{|E[B(f_1, f_2)]|^2}{E[|X(f_1)X^*(f_1)||X(f_2)X^*(f_2)||E[|X(f_1 + f_2)X^*(f_1 + f_2)|]} \\ & \equiv \frac{|E[B(f_1, f_2)]|^2}{E[|P(f_1)||P(f_2)||E[|P(f_1 + f_2)|]} \end{aligned} \quad (10)$$

Substituting the expression from equation 8 and 9, it can be shown that

$$bic^2(f_1, f_2) = \mu_3^2 / 2\pi\sigma_e^6 \quad (11)$$

From this equation, it can be seen that the squared bicoherence is constant for a linear process. It is also independent of frequency. So, if the bicoherence is not a non-zero constant in the principal domain of the bispectrum, it is concluded that the squared bicoherence is also zero. In that case, the signal is Gaussian.

3. PROPOSED WORK

19 western musical instruments with full pitch range from Mc Gill Master Sample Database used for classified by preprocessing musical signal followed by feature extraction then KNN classifier will give instruments family and instrument category. Figure 1 shows block diagram of proposed work. MFCC and HOS features are extracted from musical instruments .wav file which is recorded under stereo condition. Hierarchical classification is done by KNN classifier

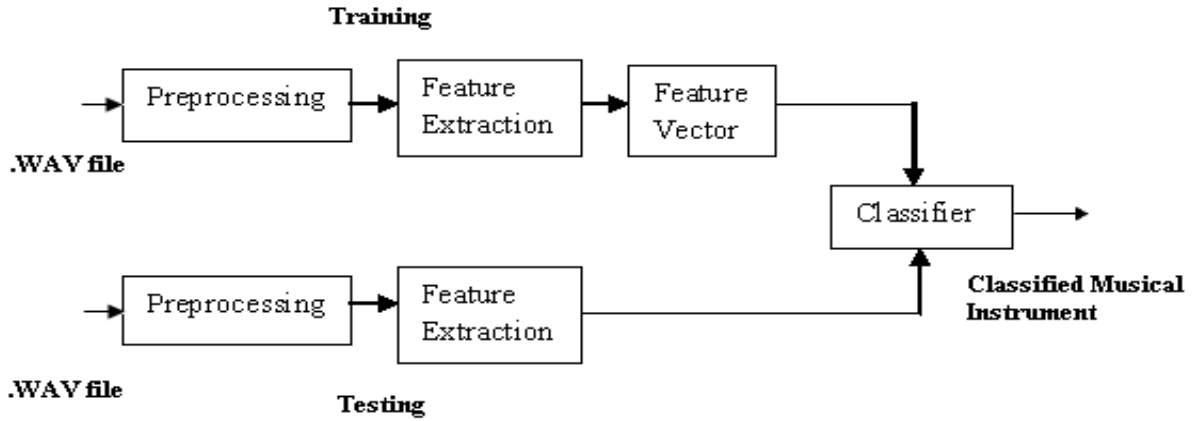


Fig 1: Block diagram of proposed work

Instruments used for proposed work are tabulated in Table 1.

Table 1. Database Used

Family	Instruments
Brass	Cornet, French horn, Trombone, Trumpet, Tuba
Percussion	Steel Drum, Tympani
String	Guitar, Bass, Cello, Harpsichord, Lute, Piano, Viola, Violin
Woodwind	Saxophone, Oboe, Oboe classical, English horn

3.1 Features Extracted

Features are numerical values extracted from signal used as input for classifier. In proposed method MFCC and HOS feature are extracted

Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are cepstral coefficients used for representing audio in a way that mimics the physiological properties of the human auditory system. MFCCs are commonly used in speech recognition and are finding increased use in music information recognition and genre classification systems. The cepstrum of a signal is the Fourier transform of the logarithm (decibel) signal of the Fourier transform of a signal. In the Mel frequency cepstrum, the frequencies are scaled logarithmically using the Mel scale. A mel is a psychoacoustic unit of frequency which relates to human perception, the mel scale can be approximated from a Hz value by the formula (12)

$$Melfrequency = 2595 \times \log_{10} \left(\frac{1+x}{700} \right) \quad (12)$$

Where x is frequency in Hz.

MFCC coefficients are amplitudes of short time power spectrum of a signal which is calculated over a non linear mel scale frequency. Power spectrum is unable to retain phase information after transformation of domain so higher order spectrum must be used to get phase information.

Bispectrum

Mean and variance will not provide more information about nonlinear signal so HOS must be used to get supplementary information. HOS suppress Gaussian noise, reconstruct phase as well as magnitude response of signal, detect & characterize nonlinearities in data. Bispectrum is Fourier transform of third

order cummulant which is estimated by using equation no (4). Li Sheng-jun and Liu Yi extracted HOS features for lung sounds [15] such as normalized bispectral entropy, Mean bispectrum magnitude, phase entropy and many more. In present work bispectrum is estimated by direct FFT method. Normalized bispectral entropy, Mean bispectrum magnitude, bispectrum phase entropy and statistics for non-gaussianity and nonlinearity are used as HOS features.

Mean Bispectrum Magnitude (mAmP):

The mean magnitude of the bispectrum can be useful in discriminating between processes with similar power spectra but different third order statistics. However, it is sensitive to amplitude changes. Normalization can easily take care of such variation. Mean bispectrum magnitude can be calculated using equation no 13.

$$mAmP = \frac{1}{L} \sum \Omega |B f1, f2| \quad (13)$$

Where $(f1, f2)$ is the bispectrum of signal.

The phase entropy would be zero if the process were harmonic and perfectly periodic and predictable. As the process becomes more random, the entropy increases. Unlike, Fourier phase, the bispectral phase does not change with a time shift. The normalization in the equations 14 and 15 ensures that entropy is calculated for a parameter that lies between 0 and 1 as required of a probability and hence the entropies ($EntP1$ and $entPh$) computed are also between 0 and

Normalized bispectral entropy (Ent1):

$$Ent1 = - \sum_n P_n \log P_n \quad (14)$$

$$\text{Where } P_n = \frac{|B(f1, f2)|}{\sum |B(f1, f2)|}$$

Bispectrum phase entropy (entPh):

$$entPh = \sum_n p\Psi_n \log \Psi_n \quad (15)$$

Where $p\Psi_n = \frac{1}{L} \sum \Omega 1\Phi b f1, f2 \in \Psi_n$ and

$$\Psi_n = \frac{\Phi}{-\pi} + \frac{2\Pi n}{N} \leq \Phi < -\Pi + \frac{2\Pi n}{N}, n = 0, 1, \dots, N-1$$

Bicoherence is a normalized bispectrum, i.e., the third order correlation of three harmonically related Fourier frequencies of a signal. Bicoherence is insensitive to signal gaussianity and bispectrum is often used as a measure of signal non-gaussianity. For non-gaussianity and non-linearity parameters chi-squared value, mean bicoherence and theoretical and estimated value for inter quartile distance are estimated using MATLAB Higher order spectral analysis toolbox.

Elena Rusticelli, Richard A.Ashley, Estela bee Dagum and Douglas M.Patterson [5] discussed the estimator $2|\hat{X}(f_m, f_n)|^2$, referred as estimated standardized bispectrum, is asymptotically non central chi-square distribution with two degrees of freedom, $\chi_{2^2}(\lambda_{m,n})$, where

$$\hat{X}(f_m, f_n) = \frac{\widehat{B}_x(f_m, f_n)}{\sqrt{\frac{N Q_{m,n}}{M^4} [\widehat{S}_x(f_m) \widehat{S}_x(f_n) \widehat{S}_x(f_{m+n})]}} \quad (16)$$

is asymptotically a standard normal and $\widehat{S}_x(\cdot)$ is a consistent estimator of the power spectrum and $f_j = (2j - 1)M/2N$ for any integer j. The non centrality parameter $\lambda_{m,n}$ is given by

$$\lambda_{m,n} = 2 \left(\frac{N Q_{m,n}}{M^4} \right)^{-1} \frac{|\widehat{B}_x(f_m, f_n)|^2}{S_x(f_m) S_x(f_n) S_x(f_{m+n})} \quad (17)$$

where $Q_{m,n}$ is equal to the number of all pairs (j, k) within the lattice square inside the triangular principal domain, but not on the boundaries of the triangle $j=k$, or $2j+k=N$. If whole square is inside the domain, then $Q_{m,n} = M^2$, otherwise $Q_{m,n} < M^2$. Hence, the value of the non centrality parameter increases when a smaller set of frequency pairs (m, n) is considered as distribution of the estimated standardized bispectrum chi-squared in order to evaluate the null hypothesis of linearity. When given data time series is linear, the parameter $\lambda_{m,n}$ is constant for all P pairs (m, n) considered in the principal domain.

Figure 2, 3& 4 shows three dimensional bicoherence plot for Oboe, Tuba and English horn musical instrument signals respectively. Sharp peaks in plot indicate nonlinearity present in the signal.

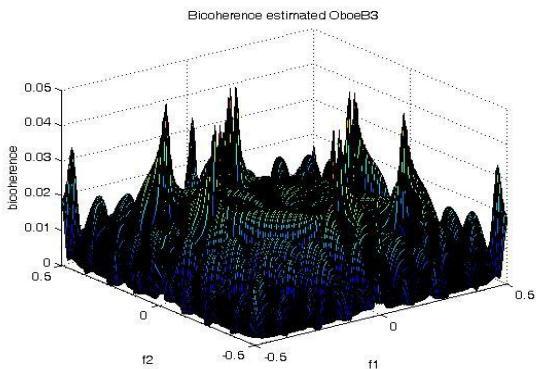


Fig 2: Bicoherence 3-D plot for Oboe

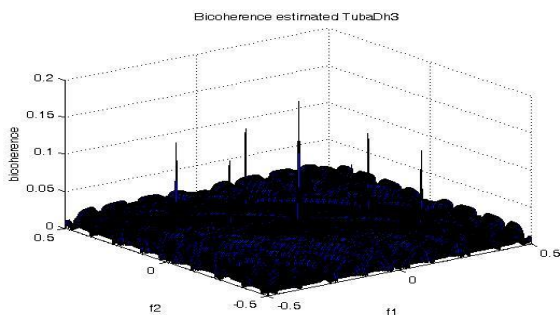


Fig 3: Bicoherence 3-D plot for Tuba

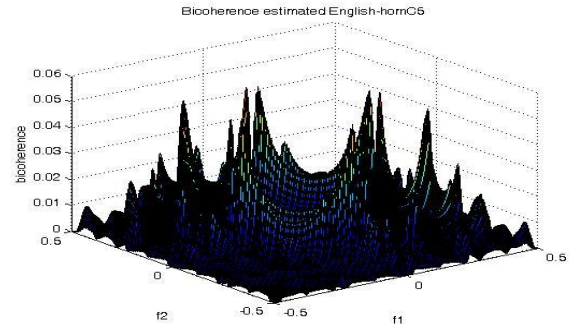


Fig 4: Bicoherence 3-D plot for English horn

3.2 Taxonomy

In 4th century BC China introduced instrument taxonomy where instruments are classified according to material from which they were made after that several musicologist's brought up various classification schemes such as the popular Sachs-Hornbostel system and grouping by playing methods have all developed for a long time. The essence of taxonomy for all science is logicity and standard, but none of them is the standard recognized all over the world or meets all application needs, which makes it hard to find a single classification scheme to satisfy all users when it comes to programs based on taxonomy. Thus musical instrument classification now has become not only a subject for musicologists, but also an important research field for Music Information Retrieval (MIR). Taxonomies improve usability, search success rate and user satisfaction. Taxonomy 1 classifies instruments into four families according to how they set air into vibrations. Taxonomy 2 has five families according to vibrating material that produces the sound. Figure 5 & 6 shows taxonomies used for present work.

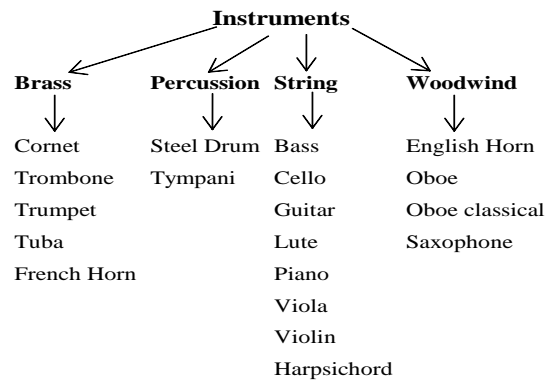


Fig 5: Taxonomy 1

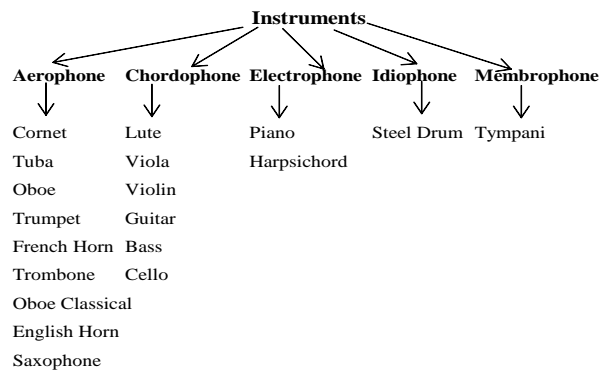


Fig 6: Taxonomy 2

4. CLASSIFIER

Jing Liu, Lingyun Xie [7] classified 13 western and 13 Chinese musical instruments using SVM with MFCC features but it does not give better result for percussion family of both western and Chinese instruments. Peter Somerville and Alexandra L. Uitdenbogerd [2] concluded MFCC with KNN classifier for polyphonic sound gives 80% classification accuracy. The K-Nearest Neighbor (K-NN) algorithm measures the distance between a query scenario and a set of scenarios in the data set. It works based on minimum distance from the query instance to the training samples to determine the K-nearest neighbors. After we gather K nearest neighbors, we take simple majority of these K-nearest neighbors to be the prediction of the query instance. KNN algorithm is described below.

1. Let X_i represent several multivariate attributes name that will be used to classify Y.

2. Calculate the distance between the query-instance and all the training samples. Because we use only quantitative X_i , we can use Euclidean distance. Suppose the query instance have coordinates (x_1^q, x_2^q) and the coordinate of training sample is (x_1^t, x_2^t) then square Euclidean distance is $d_{iq}^2 = (x_1^t - x_1^q)^2 + (x_2^t - x_2^q)^2$

3. Sort the distance of all training samples to the query instance and determine the K-th minimum distance. If the distance of the training sample is below the K-th minimum, then we gather the category Y of this nearest neighbors training samples.

70% notes for training and 30% notes are given for testing purpose. Waikato Environment for knowledge analysis (WEKA) tool [13] and MATLAB have been used for developing the algorithms.

5. RESULT

Overall classification results are obtained for both Taxonomies using MFCC and MFCC along with higher order spectra (HOS) features. Classification accuracy got improved when HOS features are used with MFCC. Figure 7 & 8 shows graph of classification result for Taxonomy 1 and Taxonomy 2 respectively.

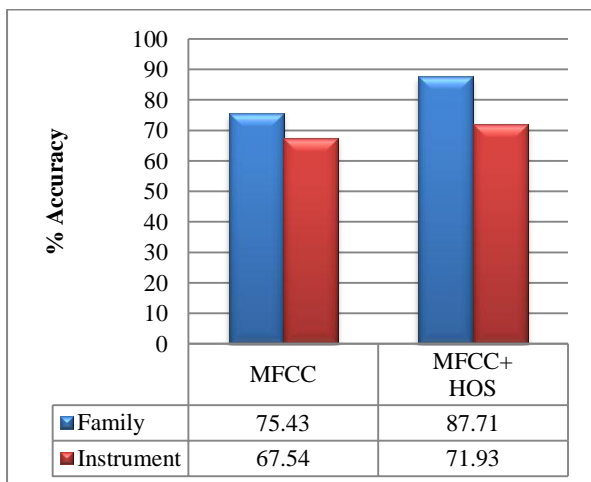


Fig 7: Classification accuracy graph for Taxonomy 1

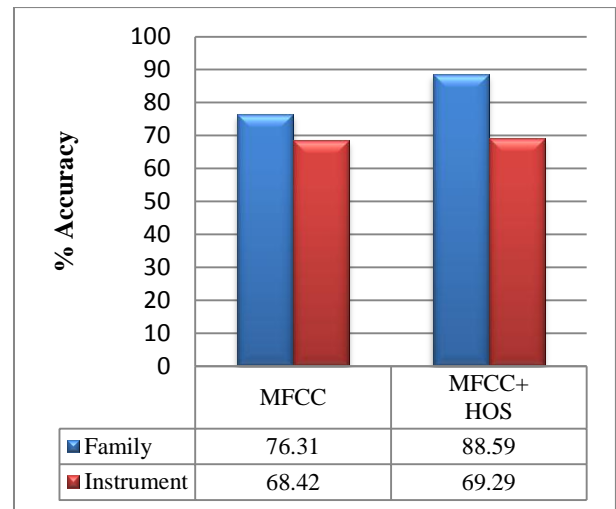


Fig 8: Classification accuracy graph for Taxonomy 2

6. CONCLUSION

Higher order spectra provide the additional information regarding phase entropy, non gaussianity and non linearity which helps to provide more information to the classifier. With addition of higher order spectra features to MFCC features accuracy of classification increases. Both taxonomies show improved result with MFCC+HOS. Taxonomy 2 gives better classification accuracy than taxonomy 1 for instrument in respective families

Different taxonomies can be implemented to classify musical instruments. Feature selection algorithm can be used at different hierarchical level to improve classification accuracy.

7. REFERENCES

- [1] Glenn Eric Hall, Hassan Ezzaidi, Mohammed Bahoura "Hierarchical Parameterization and Classification For Musical Instrument Recognition", 11th International Conference On Information Sciences, Signal Processing and their Applications Main Tracks, PP.1039-1044, 2012
- [2] Peter Somerville and Alexandra L. Uitdenbogerd.: "Multitimbral Musical Instrument Classification", International Symposium on Computer Science and its Applications PP.269-274, 2008.
- [3] Saima Anwar Lashari, Rosziati Ibrahim and Norhalina Senan "Soft Set Theory for Automatic Classification of Traditional Pakistani Musical Instruments Sounds" International Conference on Computer & Information Science (ICIS), pp.94-99, 2012
- [4] Rui Rui and Changchun Bao. "A Novel Supervised Learning Algorithm for Musical Instrument Classification" 35th International conference on Telecommunication and Signal Processing ,pp. 446-449, 2012
- [5] Elena Rusticelli, Richard A. Ashley, Estela bee Dagum and Douglas M. Patterson "A New Bispectral Test for Nonlinear Serial Dependence" Taylor & Francis Journals, vol.28(1-3), 279-293,
- [6] Akram Azarloo, Fardad Farokhi.: Automatic Musical Instrument Recognition Using K-NN and MLP Neural Networks" in 4th International Conference on Computational Intelligence, Communication Systems and Networks, pp. 289-294, 2012.

- [7] Jing Liu, Lingyun Xie. "SVM-Based Automatic Classification of Musical Instruments" International Conference on Intelligent Computation Technology and Automation, pp.669-673, 2012.
- [8] S.Gunasekaran, K.Revathy "Fractal Dimension Analysis of Audio Signals for Indian Musical Instrument Recognition" International conference on Audio, language & image processing ICALIP, pp.257-261,2008
- [9] Alicja A. Wiczorkowska, Zbigniew W. Ra's, Xin Zhang, Rory Lewis "Multi-way Hierarchic Classification of Musical Instrument Sounds" International Conference on Multimedia and Ubiquitous Engineering, pp.897-902, 2007
- [10] Roisin Loughran, Jacqueline Walker, Michael O'Neill, Marion O'Farrell "Musical Instrument Identification Using Principal Component Analysis & Multi-Layered Perceptrons" International conference on Audio ,Language & Image processing ICALIP,pp.643-648, 2008.
- [11] Slim ESSID, Ga'El Richard & Bertrand David "Hierarchical Classification of Musical Instruments on Solo Recordings" IEEE international conference on Acoustic speech & signal processing ICASSP proceeding, pp.68-80,2006
- [12] McGill University Master Sample www.music.mcgill.ca/resources/mum/.html/mums.html
- [13] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten The WEKA Data Mining Software: an Update; SIGKDD Explorations, Volume 11, Issue 1, (2009)
- [14] M.A.A.S. Choudhury, S. L. Shah, N. F. Thornhill "Detection and diagnosis of System Nonlinearities using higher order statistics" 15th Triennial world congress ,Barcelona, Spain, pp. 1-6, 2002.
- [15] LI Sheng-jun, LIU Yi "Feature Extraction of Lung Sounds Based on Bispectrum Analysis" Third International Symposium on Information processing pp. 393-397, 2010.