www.czasopisma.pan.pl PAN www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **41**, No. 3, pp. 427–436 (2016)

DOI: 10.1515/aoa-2016-0042

Research Papers

Hybridisation of Mel Frequency Cepstral Coefficient and Higher Order Spectral Features for Musical Instruments Classification

Daulappa Guranna BHALKE[(1)], C.B. RAMA RAO[(1)], Dattatraya BORMANE[(2)]

[(1)] *National Institute of Technology*
Warangal, India; e-mail: bhalkedg2000@yahoo.co.in, cbrr@nitw.ac.in

[(2)] *Rajarshi Shahu College of Engineering*
Tathawade, Pune, India; e-mail: bdattatraya@yahoo.com

(*received November 13, 2014; accepted February 5, 2016*)

This paper presents the classification of musical instruments using Mel Frequency Cepstral Coefficients (MFCC) and Higher Order Spectral features. MFCC, cepstral, temporal, spectral, and timbral features have been widely used in the task of musical instrument classification. As music sound signal is generated using non-linear dynamics, non-linearity and non-Gaussianity of the musical instruments are important features which have not been considered in the past. In this paper, hybridisation of MFCC and Higher Order Spectral (HOS) based features have been used in the task of musical instrument classification. HOS-based features have been used to provide instrument specific information such as non-Gaussianity and non-linearity of the musical instruments. The extracted features have been presented to Counter Propagation Neural Network (CPNN) to identify the instruments and their family. For experimentation, isolated sounds of 19 musical instruments have been used from McGill University Master Sample (MUMS) sound database. The proposed features show the significant improvement in the classification accuracy of the system.

Keywords: feature extraction; MFCC; HOS; bispectrum; bicoherence; non-linearity; non-Gaussianity; CPNN; Zero Crossing Rate (ZCR).

1. Introduction

The human ability to distinguish among musical instrument sounds has been a subject of investigation for the last few decades. Even with minimal musical knowledge exposure, most people can easily distinguish among familiar musical instruments, even when they are played at the same loudness and pitch. Timbre is the quality of sound by which a listener can distinguish between two sounds of equal loudness, duration and pitch, but it has been proven to be difficult to measure or quantify. In addition, the demand for online access to music data in the Internet is increasing day by day.

The objective of this paper is to present the importance of HOS-based features in the task of musical instrument classification. HOS-based features provide information related to non-linearity and non-Gaussianity of musical instruments. This additional information helps to improve the overall classification accuracy of the system.

Musical instrument classification problem consists of three steps: pre-processing, feature extraction, and classification. The majority of research on musical instrument classification is focused on feature extraction. The state-of-the art on musical instrument classification has been described below.

Kostek (Kostek, Wieczorkowska, 1997; Kostek, Krolikowski, 1997; Kostek, Czyzewski, 2001) has presented effectiveness of spectral and temporal features for musical instrument identification and classification. Also, she has demonstrated musical instrument sound classification using a limited number of parameters and neural network. An expert system was built for automatic musical instrument classification using a rough set of parameters with a fuzzy based approach (Kostek, 2004a). Further, the use of soft computing techniques in the field of music acoustics was fully justified for musical sound classification (Kostek, 2004b; 2007). The performance of different classifiers on the selected feature set has also been evaluated using Waikato Environment for Knowledge analysis (WEKA) (Kostek, Kania, 2008). Finally she concluded that finding novel features for musical instrument classification is a major challenge and future scope of research. Martin and Kim (1998) demonstrated the utility of hierarchical organisation of musical sounds. Log-lag correlogram acoustic features were extracted to classify 15 musical instruments. The performance of 90% accuracy was reported for families of instruments and 70% accuracy was reported for individual instruments. However, in all taxonomies string and brass family instruments have shown consistent results, but woodwind instruments were inconsistent. Eronen (2001) used features including MFCC, delta MFCC, LPCC, temporal, spectral, and modulation features for instrument classification. MFCC features have shown good performance among all. Large variations in recognition accuracies of different instruments have also been seen. Agostini *et al.* (2001; 2003) discussed content based classification of musical timbres. Different pattern recognition techniques such as support vector machine (SVM), KNN, canonical and quadratic discriminant analysis have been used to classify the musical instrument timbres. Also, Kaminskyj and Czaszejko (2005) described recognition of isolated monophonic musical sounds using KNN classifier. Cepstral coefficients, amplitude envelope, constant Q-transform, multidimensional scaling analysis trajectories, spectral centroid, and vibrato features have been used for recognition of musical sounds. An accuracy of 93% for individual instruments and 97% for instrument family classification was reported for 19 instruments. Consequently, Essid *et al.* (2006) presented a work on use of natural and instrument hierarchical taxonomies for recognition of musical instruments on solo recordings. A wide set of features covering temporal, cepstral, spectral, wavelet, and perceptual features were used to classify instruments using a SVM classifier with selected features. Cepstral, spectral, perceptual, and MFCC features were included into the selected feature subset in the proposed taxonomy. The instrument based taxonomy has shown significant improvement using the selected feature subset. Further, Eronen, and Klapuri (2000) presented a study on musical instrument classification using a wide set of temporal and spectral features. The usefulness of the hierarchical structured classifier has also been demonstrated. The authors concluded that combining different type of features improved the classification accuracy of the system. Further, Loughran *et al.* (2008) described a work on musical instrument classification using MFCC features and principal component analysis. Multilayer perceptron was used as a classifier to test the performance of the system. Optimum numbers of coefficients were determined to classify musical instrument samples of piano, violin, and flute. Deng *et al.* (2008) reviewed past research work on instrument classification and described feature analysis for musical instrument classification using different machine learning techniques. A large set of features including MPEG-7, perceptual based features, statistical features of MFCC, statistical features of ZCR, RMS energy, spectral centroid, and spectral flux were used to form feature subset for instrument recognition. MFCC features were found dominating among the selected features subset.

In addition, literature review on HOS analysis for signal processing has been carried out and briefly described. Dubnov and Tishby (1994; 1997) demonstrated the use of the higher order statistics of acoustic signals for different spectral estimation and modelling various auditory perceptual phenomena. Sound textures and machine sound has also been analysed using HOS-based features. It has been shown that the higher order statistics captures timbral information better and provides additional information about musical instrument sound signals. Again, Dubnov and Tishby (1998) have shown that the non linear properties in musical sound signals are attributed to excitation source of musical instruments, and spectral properties are attributed to resonating chambers of musical instruments. Skewness and kurtosis features have been used to characterise the residual part of musical sound signal and exhibit similar results using higher order spectra based features. Choudhury *et al.* (2002) used the higher order statistics to detect and quantify the non-Gaussianity and non-linearity of regulated processes or control error variables which were the main contributors to the poor performance of many of the control loop. Bicoherence plots and bicoherence index were used to detect non-linearities. Li and Liu (2010) extracted features from lung sounds of normal, pneumonia, and asthma patients in bi-frequency domain. Further, Dubnov and Rodet (2005) demonstrated that phase coupling is an important characteristic of a sustained portion of sound of individual musical instruments, and effect of phase coupling has been compared by means of higher order statistics. In addition, the timbre of pitched musical instrument was analysed by Liu *et al.* (2010) using excitation signature by means of the higher order statistics and subspace analysis. It has been demonstrated that HOS-based features provide more significant timbre patterns in both time and frequency domains in comparison with the second order statistics. Ajmera *et al.* (2012) extracted MFCC features from bispectrum which have been used to reconstruct the spectrum of the original signal from its noisy version and observed improvement in recognition accuracy. Goshvarpour *et al.* (2012) used the bispectrum for EEG signal analysis. Bordoloi *et al.* (2012) used hybrid features of bispectrum for classification of Motor imagery (MI). Also, Bhalke *et al.* (2014) have demonstrated the significance of the bispectrum and bicoherence in musical instrument classification. In general the literature review can be summarised as follows: various feature schemes and classification algorithms have been proposed for musical instrument classification. However, MFCC features were found dominating for musical instrument classification. It has also been seen that finding new acoustic features is a challenging and future scope of the work. Further, HOS-based features have been utilised in many applications to provide information about non-linearity, non-Gaussianity, and phase related information of musical sound signals, but they have not been used for musical instrument classification. In this paper, an attempt has been made to improve the classification accuracy of the system using hybridisation of HOS-based features and MFCC features.

Following the introduction and state-of-the-art of work, Sec. 2 outlines the methodology used in tackling the problem of instrument classification including feature extraction and classification algorithms. The experimental settings and results based on the proposed approach are presented in Sec. 3. Finally, conclusions are presented in Sec. 4.

2. Methodology

The proposed methodology for instrument classification has been depicted in Fig. 1. It consists of training and testing phase. In the training phase, initially a signal is pre-processed to remove the silence part of a signal using energy and ZCR features. It helps to reduce the computational complexity of the system. After pre-processing, suitable features like MFCC and HOS-based features are extracted. The extracted features have been presented to CPNN to build classification model for each feature scheme. In the testing phase, a signal is again pre-processed and similar features as in the training are extracted. The extracted features are then presented to CPNN to make a decision. The feature extraction and experimental details have been presented in Subsecs. 2.1 and 2.2, respectively.

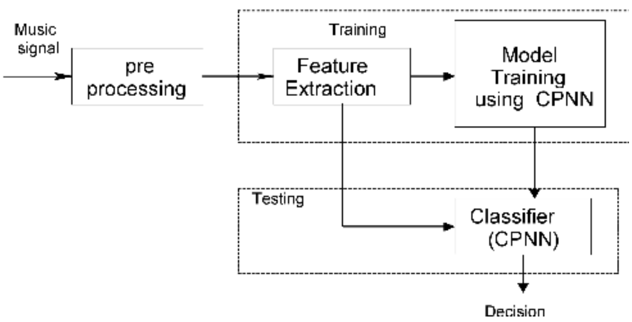

Fig. 1. Musical instrument classification system.

Database used:

McGill University Master Samples (MUMS) musical sound database has been used for the experiments. The database has been created by Frank Opolko and Joel Wapnick (1987). The music samples were recorded at 44.1 KHz with a wide variation of articulation styles and pitch ranges. Nineteen instruments from four families have been used for experimentation. The list of the instruments used for experimentation has been given in Table 1. These instruments have been widely used in various research studies as discussed in the literature review. Forty isolated notes of each instrument have been used for experimentation.

Table 1. List of the musical instruments.

Sr. No.	Instruments	Family
01	Trumpet, Cornet, Tuba, Trombone, French Horn	Brass
02	Saxophone, Oboe classical, Oboe D, English horn	Woodwind
03	Violin, Viola, Guitar, Cello, Lute, Piano, Harpsichord, Bass	String
04	Steel drum, Tympani	Percussion

2.1. Feature extraction

Feature extraction is used to obtain the relevant and significant information about a signal. Features including MFCC and HOS-based features have been extracted. The MFCC feature extraction and HOS-based feature extraction have been briefly presented in Subsecs. 2.1.1 and 2.1.2, respectively. Table 2 shows the features used for experimentation.

Table 2. List of features.

Feature No.	Name of the feature
1–13	Mean value of MFCC coefficients
14–20	Bispectrum features: normalized bispectral entropy, bispectrum phase entropy, normalized bispectral squared entropy, normalized bispectral cubed entropy, mean bispectrum magnitude, and mean value of bicoherence

2.1.1. Mel Frequency Cepstral Coefficients (MFCC)

The MFCC features have proved their significance in speech recognition, speaker recognition, and also in musical instrument classification, Eronen (Eronen, 2000; Eronen, Klapuri, 2001), Martin and Kim (1998), Deng *et al.* (2008), Loughran *et al.* (2008), Bhalke *et al.* (2015). MFCC represents the short time power spectral representation of a sound signal. It provides useful information regarding psychoacoustic properties of human auditory system. The block

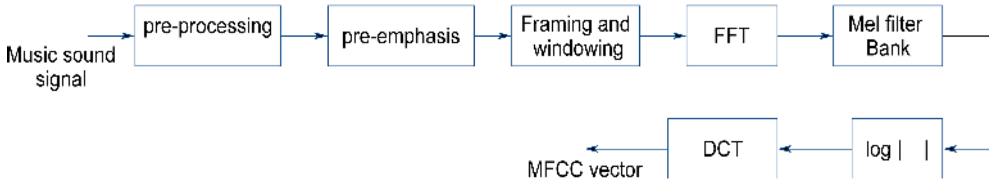

Fig. 2. MFCC Feature extraction.

scheme of MFCC has been depicted in Fig. 2. It consists of pre-processing, pre-emphasis, framing, windowing, FFT, Triangular mel band pass filter, and DCT. In pre-processing the silence part of a signal has been removed using ZCR and energy features with a proper threshold value. Pre-emphasis has been done to boost the high frequency components of a signal using the first order high pass filter. Framing has been done with 20 ms duration with 10 ms overlap. After framing a signal, it is windowed using a Hamming window. Further, the signal is transformed into spectral domain and passed through 24 mel frequency triangular band pass filters. Log values of these spectral components have been obtained. Discrete Cosine Transform (DCT) of these log values have been taken to decorrelate the features. Thirteen most significant MFCC coefficients have been obtained for each frame. Mean value of these coefficients have been computed and used as a feature vector.

2.1.2. Higher Order Spectral Analysis (HOSA)

First and second order statistics such as mean, variance, autocorrelation, power spectral density are frequently used signal processing tools. They are useful for linear and Gaussian processes, but when data are deviated from Gaussianity and linearity these tools were found to cause shortcomings for analysis. HOS is mainly used to extract information due to deviation from Gaussianity, to recover phase information, and to detect nonlinearities. As music sound signals are generated using non linear dynamic process, HOS-based features help to provide additional information about musical instruments. These non linear dynamic features have been extracted using bispectrum and bicoherence. The following sections briefly describe the bispectrum and bicoherence based features.

Bispectrum:

Power spectrum is the frequency domain representation of the second order moment but it does not provide information about the higher order moment. The bispectrum is the frequency domain representation of the third order cumulants. It is Fourier transform of the third order cummulant and is depicted in Eq. (1).

$$B(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)], \tag{1}$$

where $X(f_1)$ is Fourier transform of a signal $x(t)$. Bispectrum is a complex quantity having both magnitude and phase. It is plotted against two independent frequency variables f_1 and f_2 in three-dimensional plots. Each point in the plot represents bispectral content of a signal at the bifrequency $(f_1,\ f_2)$. In this paper, bispectrum has been computed using direct FFT method. Features such as normalised bispectral entropy, mean bispectrum magnitude, bispectrum phase entropy have been estimated based on the bispectrum. Figure 3 shows bispectrum plot of Trumpet C6 notes and Fig. 4 shows bispectrum plot for Tuba C3 note. Figures 3 and 4 show that the bispectrum plots are different for different instruments. The features extracted from the bispectrum help to distinguish the instruments.

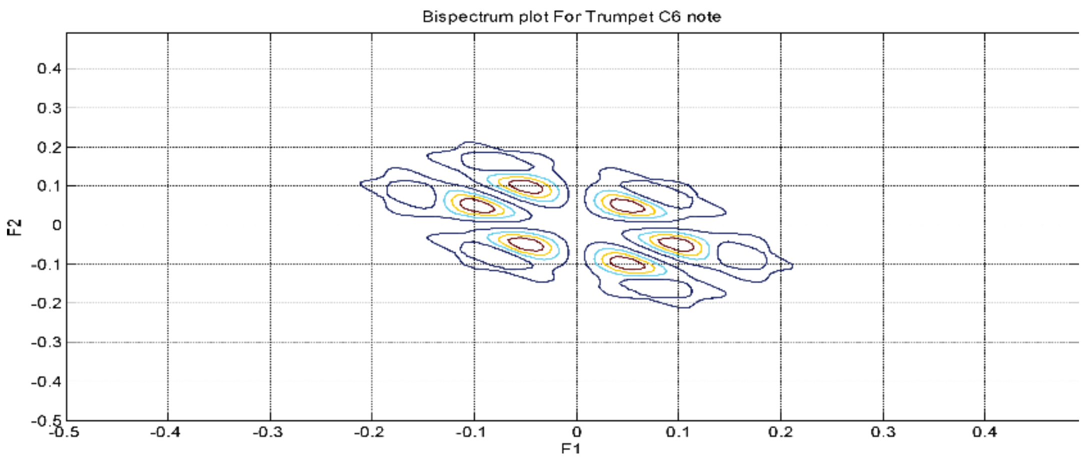

Fig. 3. Bispectrum plot for Trumpet C6 note.

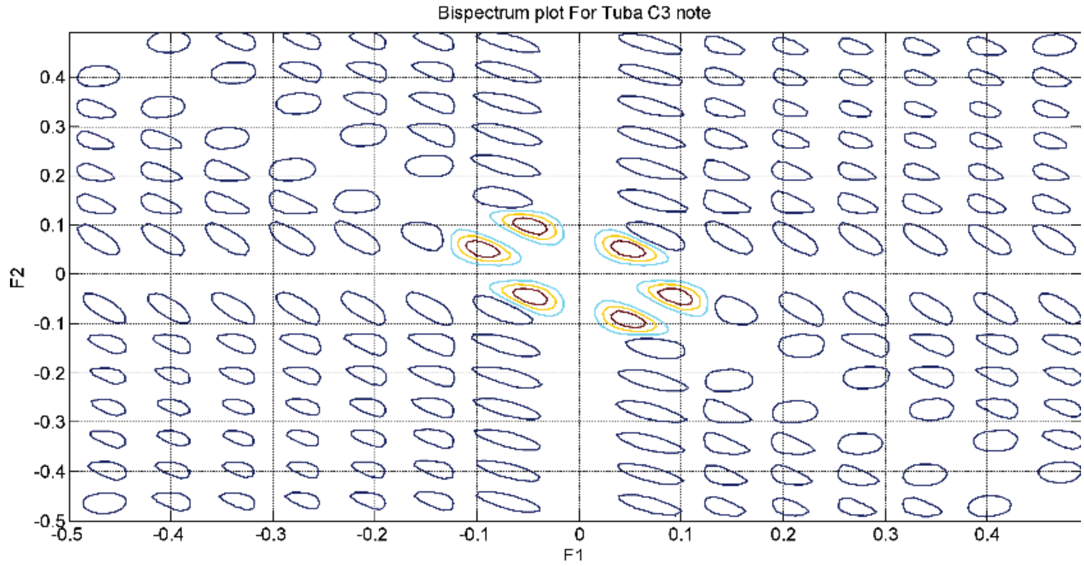

Fig. 4. Bispectrum plot Tuba C3 note.

The following features have been extracted to measure the similarity of the bispectrum.

Normalised Bispectral entropy (Ent1):

Normalised bispectral entropy is computed using Eq. (2).

$$\begin{aligned} Ent1 &= -\sum_n p_n \log p_n, \\ \text{where} \quad p_n &= \frac{|B(f_1, f_2)|}{\sum_\Omega |B(f_1, f_2)|}, \end{aligned} \tag{2}$$

where $B(f_1, f_2)$ is the bispectrum of a signal.

Normalised bispectral squared entropy (Ent2):

Normalised bispectral squared entropy is computed using Eq. (3).

$$\begin{aligned} Ent2 &= -\sum_n q_n \log q_n, \\ \text{where} \quad q_n &= \frac{|B(f_1, f_2)|^2}{\sum_\Omega |B(f_1, f_2)|^2}. \end{aligned} \tag{3}$$

Normalised bispectral cubed entropy (Ent3):

Normalised bispectral cube entropy is computed using Eq. (4).

$$\begin{aligned} Ent3 &= -\sum_n r_n \log r_n \\ \text{where} \quad r_n &= \frac{|B(f_1, f_2)|^3}{\sum_\Omega |B(f_1, f_2)|^3}. \end{aligned} \tag{4}$$

Bispectrum phase entropy (EntPh):

Normalised bispectral phase entropy is computed using Eq. (5).

$$EntPh = \sum_n p(\Psi_n) \log p(\Psi_n), \tag{5}$$

where $p(\Psi_n) = \frac{1}{L}\sum_\Omega 1(\Phi(B(f_1, f_2) \in \Psi_n))$, $-\pi+2\pi n/N \le \Phi < -\pi + 2\pi(n+1)/N$, $n = 0, 1, \cdots N-1$, Φ – bispectrum phase angle, L – number of points within the samples, $1(.)$ – indicator function.

Mean bispectrum magnitude (mAmp):

The mean value of bispectrum magnitude is useful in discriminating between processes with similar power spectra but different third order statistics. Normalisation has been done to improve the scalability of features. Mean bispectrum magnitude has been computed using Eq. (6).

$$mAmp = \frac{1}{L}\sum \Omega |B(f_1, f_2)|, \tag{6}$$

where $B(f_1, f_2)$ is the bispectrum of a signal.

If the processes were harmonic, periodic then phase entropy would be zero and if they become more random, the entropy increases. Unlike Fourier phase, the bispectral phase does not change with a time shift.

Bicoherence:

Bicoherence is a normalised bispectrum of a signal which is the third order correlation of three harmonically related Fourier frequencies. It is insensitive to signal Gaussianity. The squared bicoherence is computed using Eq. (7).

$$bich^2(f_1, f_2) = \frac{|E[B(f_1, f_2)]|^2}{P(f_1)P(f_2)P(f_1, f_2)}, \tag{7}$$

where '*bich*' is bicoherence function and $P(f)$ are power spectra of the signal.

Linearity test:

Bicoherence has been used to test the linearity of a signal, Chaudhari *et al.* (2002). The following section briefly describes the linearity test using the bicoherence value.

Let $y(k)$ be a discrete stationary process. The linear system can be represented by Eq. (8):

$$y(k) = \sum\nolimits_n h(n)x(k-n), \tag{8}$$

where $x(k)$ is a sequence of independent identically distributed random variables with

$$E[x(k)] = 0, \quad \sigma_x^2 = E[x^2(k)], \quad \text{and} \quad \mu_3 = E[x^3(k)].$$

The power spectrum for this linear system is represented by Eq. (9),

$$P(f) = \frac{\sigma_x^2}{2\pi}|H(f)|^2 \tag{9}$$

and the bispectrum is given by

$$B(f_1, f_2) = \frac{\mu_3}{(2\pi^2)}H(f_1)H(f_2)H^*(f_1+f_2), \tag{10}$$

where $H(f)$ is Fourier transform of $h(n)$. Bicoherence is normalised bispectrum of a signal and the squared bichorence is given by Eq. (7) which is

$$bich^2(f_1,f_2) = \frac{|E[B(f_1,f_2)]|^2}{P(f_1)P(f_2)P(f_1,f_2)}.$$

Putting the values of Eq. (9) and Eq. (10) into Eq. (7) the $bich^2(f_1, f_2)$ can be written as

$$bich^2(f_1,f_2) = \frac{\mu_3}{2\pi\sigma_x^6}. \tag{11}$$

From Eq. (11), it can be concluded that the squared bicoherence is constant for a linear process and independent of frequency. Also, if it is not a non-zero constant in the principal domain of the bispectrum, the system is non linear. Further, if μ_3 is zero the squared bicoherence is zero and the signal is Gaussian.

So, non-linearity and non-Gaussianity can be estimated using the bicoherence value. If sharp peaks are present in the bicoherence plot and mean value of bicoherence is not zero then the system is non linear. The extent of non-linearity can be estimated from the value of bicoherence. More non-linearity is present if the bicoherence value is maximum and *vice versa*. Figure 5 shows the 3D bicoherence plot for Piano C2 note and Fig. 6 shows the 3D bicoherence plot for Cello C4 note. These figures show that the extent of non-linearity is different for different instruments. It can be easily estimated using the mean bicoherence value.

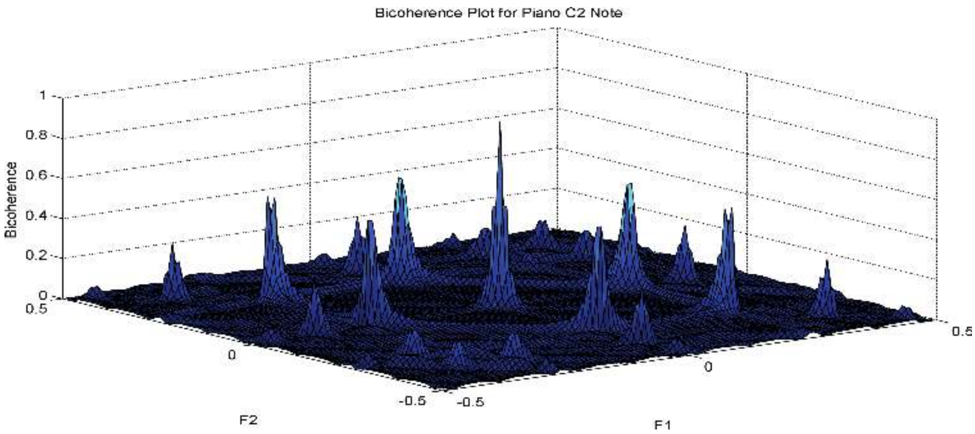

Fig. 5. 3D bicoherence plot for Piano C2 note.

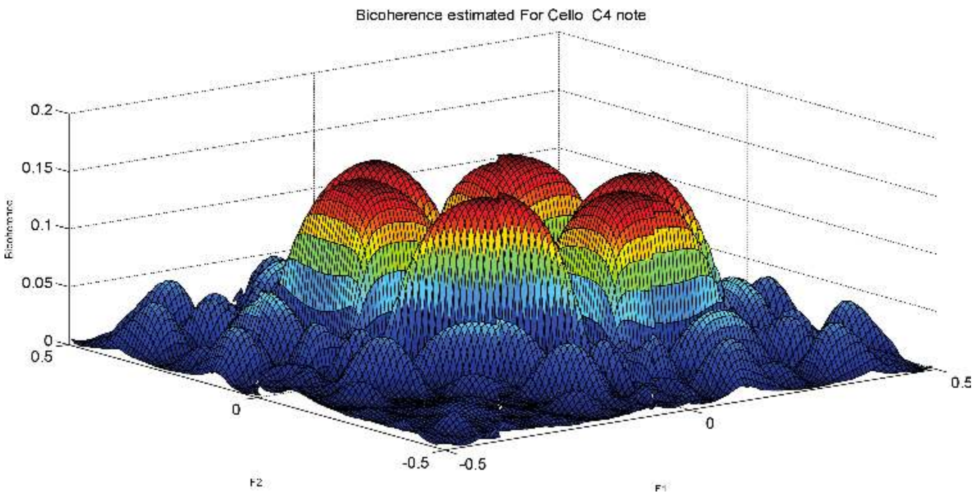

Fig. 6. Bicoherence plot for Cello C4 note.

2.2. Counter Propagation Neural Network (CPNN)

Neural Networks (NN) have been used to solve classification and forecasting problems for the last two decades. Kohonen's Self Organizing Map (SOM) is one of the most widely used NN to solve the classification problem in an unsupervised manner. It has been used as a powerful data analysis and visualisation tool in many applications. It has the ability to organise input vectors in an unsupervised learning fashion. It basically focuses on self organising and clustering applications. The extension of SOM is Counter Propagation Neural Network (CPNN). It is one of the frequently used NN to solve many classification problems in a supervised manner, Kuzmanovski and Novič (2008). CPNN has an excellent clustering ability (unsupervised) and reduces errors existing in the desired output (supervised learning). The purpose of using CPNN is to take the benefits of both the supervised and unsupervised learning. It consists of three layers: input, Kohonen's, and output one. Input data (extracted features) are introduced to the input layer. Kohonen's layer clusters the input vector using Euclidean distance and the winner takes all rule, and weights of cluster nodes are computed. Weights of cluster nodes are introduced to the output layer to get the desired target vector. The learning speed of CPNN is faster as compared to other NNs. The training model has been developed using CPNN.

The steps for CPNN training are given below:

Step 1: Apply x dimensional input vector and y dimensional output vector to the network.

Step 2: Compute the distance between the input and weight vector of Kohonen layer.

Step 3: Compute the winning neuron.

Step 4: Adjust the weights of all neurons.

Step 5: Apply the new pair of input and output vectors.

Step 6: Go to step 2 till all input has been applied.

After training the CPNN model has been developed for each feature set. In the testing phase, after the winning neuron is found (Kohonen layer) the root-mean-square-error of prediction (RMSEP) is computed using weights in the output layer and prediction is done. The following parameters of CPNN have been used during training and prediction: *Network topology: Square, Network size: 10X10*, No. of Epochs: 200, *learning rate:* 0.1, *Training: Batch.*

CPNN has been selected as a classifier because it leads to faster training, better prediction, higher ability to adaptation to complex nonlinear data, lower risk of local minima, and better stable convergence as compared to other artificial neural networks (ANNs) (Goppert, Rosenstiel, 1993; Kuzmanovski, Novič, 2008). In addition, the state-of-the-art classifier such as support vector machine has serious drawbacks such as choice of the kernel function, slow speed, and high memory requirement (Byun, Lee, 2002).

3. Results and discussion

MFCC and HOS-based features have been extracted for 19 musical instruments and presented to CPNN to build a training model. Batch training with 10-cross fold validation technique have been used to train the CPNN. 70% notes have been used for training and 30% notes have been used for testing. It has been observed that musical instrument classification has shown significant improvement in classification accuracy when MFCC features have been incorporated together with HOS-based features. The result reveals that HOS-based features provide supplementary information like non-linearity, non-Gaussianity, and phase related information about acoustics of musical instruments. Figure 7 shows non-linearity of the instruments using the mean bicoherence value. The maximum value of non-linearity has been observed for the brass family and the minimum value was found for the string family instruments. Figure 8 shows classification accuracy for MFCC features combined with HOS-based features. This shows that the classification accuracy for individual instruments and instrument family has been improved significantly due to additional information provided by HOS-based features. Classification accuracy has been improved from 75% (for MFCC) to 81.39% (for the proposed features) for individual instruments and from 78.98% (for MFCC) to 87.50% (for the proposed features) for the family classification.

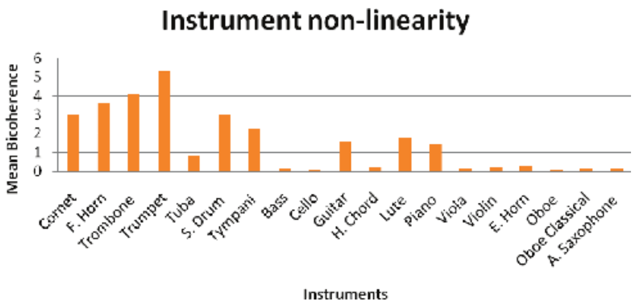

Fig. 7. Classification accuracy for different feature scheme.

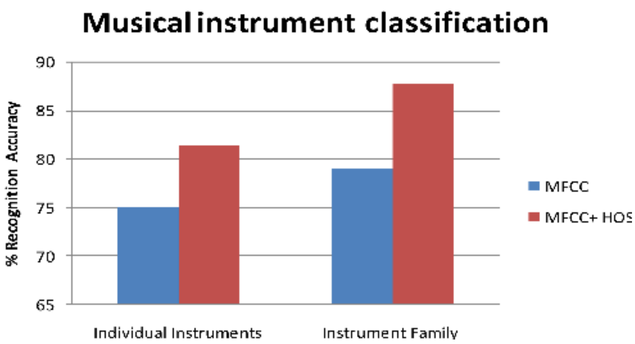

Fig. 8. Classification accuracy for different feature scheme.

Table 3 provides classification accuracy for individual instruments using MFCC features in terms of confusion matrix. The overall average classification accuracy for individual instruments reported using MFCC feature scheme is 75%. The lowest classification accuracy of 40% has been reported for the piano. Classification error in piano samples mainly occurred with guitar, lute, tympani, and drum. 20% percent samples of piano were misclassified to guitar, 15% to lute, 5% to drum, and 20% to tympani. Maximum classification accuracy of 95% has been reported for trombone. 5% of trombone samples have been misclassified to French horn. Most of the misclassification has taken place for instruments of the same family. The classification accuracy using MFCC is not significant since it provides only auditory perception information.

Table 4 provides classification accuracy for instrument families using MFCC features in terms of a confusion matrix. The overall average classification accuracy for instrument families reported using an MFCC feature scheme is 78.98 %. The lowest classification accuracy of 60% has been reported for the percussion family. Classification error in the percussion family samples mainly occurred with the woodwind family. 40% percent samples of percussion were misclassified to the woodwind family. Maximum classification accuracy of 88% has been reported for the brass family.

Table 5 provides classification accuracy for individual instruments using the proposed MFCC and HOS-based features in terms of a confusion matrix. The overall average classification accuracy for individual instruments reported using the proposed feature scheme is 81.39%. The lowest classification accuracy of 70% has been reported for lute. Classification errors in lute instrument samples mainly occurred with tympani. Maximum classification accuracy of 100% has been reported for trombone. An improvement of 6.39% has been seen because of additional information provided by HOS-based features. Table 6 provides classification accuracy for instrument families using the proposed features in terms of a confusion matrix. The overall average classification accuracy for instrument families reported using the proposed feature scheme is 87.50%. The lowest classification accuracy of 73.42% has been

Table 3. Confusion matrix for individual instruments classification using MFCC features (all numbers are in %).

Instruments	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A = Saxophone	**75**	0	10	0	5	0	0	0	0	0	0	0	10	0	0	0	0	0	0
B = Bass	5	**65**	5	0	5	0	0	0	0	0	0	5	0	0	0	0	0	15	0
C = Cello	5	10	**65**	0	0	0	0	0	0	0	0	0	0	0	1	0	5	5	5
D = Cornet	5	0	0	**90**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
E = Eng Horn	5	0	0	0	**65**	0	0	0	0	15	5	0	0	0	5	0	0	5	0
F = F_Horn	0	0	0	0	0	**90**	0	0	0	0	5	0	0	5	0	0	0	0	0
G = Guitar	0	0	0	0	0	0	**70**	0	10	0	0	0	15	0	0	0	5	0	0
H = Harpsichord	0	0	0	0	0	0	5	**75**	5	0	0	0	5	0	0	0	0	5	5
I = Lute	0	0	0	0	0	0	5	0	**60**	0	0	0	0	0	0	0	30	0	0
J = Oboe classical	0	0	0	0	15	0	0	0	0	**65**	10	0	0	0	0	0	0	5	0
K = Oboe D	0	0	0	0	15	0	0	0	0	5	**70**	0	0	0	0	0	0	5	0
L = Piano	0	0	0	0	0	0	20	0	15	0	0	**40**	5	0	0	0	20	0	0
M = Drum	0	0	0	0	0	0	10	0	10	0	0	0	**80**	0	0	0	0	0	0
N = Trombone	0	0	0	0	0	5	0	0		0	0	0	0	**95**	0	0	0	0	0
O = Trumpet	10	0	5	0	0	0	0	0	0	0	0	0	0	5	**80**	0	0	0	0
P = Tuba	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	**90**	0	0	0
Q = Tympani	0	0	0	0	0	0	0	0	15	0	0	5	0	0	0	0	**80**	0	0
R = Viola	0	5	0	0	0	0	5	0	0	0	0	5	0	0	0	0	0	**85**	0
S = Violin	0	0	5	10	0	0	0	0	0	0	0	0	0	0	0	5	0	0	**80**

Table 4. Confusion matrix for instrument family classification using MFCC features (all numbers are in %).

	String	Woodwind	Brass	Percussion
String	**83.33**	2.56	1.92	12.18
Woodwind	8.75	**78.75**	10	2.5
Brass	7	4	**88**	1
Percussion	0	40	0	**60**

Table 5. Confusion matrix for individual instruments classification using MFCC and HOS features (all numbers are in %).

Instruments	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A = Saxophone	**80**	0	10	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
B = Bass	5	**70**	5	0	0	0	0	0	0	0	0	5	0	0	0	0	0	15	0
C = Cello	5	5	**70**	0	0	0	0	0	0	0	0	0	0	0	5	0	5	5	5
D = Cornet	5	0	0	**95**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E = Eng Horn	5	0	0	0	**70**	0	0	0	0	10	5	0	0	0	5	0	0	5	0
F = F_Horn	0	0	0	0	0	**95**	0	0	0	0	0	0	0	5	0	0	0	0	0
G = Guitar	0	0	0	0	0	0	**75**	0	5	0	0	0	15	0	0	0	5	0	0
H = Harpsichord	0	0	0	0	0	0	10	**80**	5	0	0	0	0	0	0	0	0	0	5
I = Lute	0	0	0	0	0	0	5	0	**65**	0	0	0	0	0	0	0	30	0	0
J = Oboe classical	0	0	0	0	15	0	0	0	0	**70**	10	5	0	0	0	0	0	0	0
K = Oboe D	0	0	0	0	15	0	0	0	0	5	**75**	0	0	0	0	0	0	5	0
L = Piano	0	0	0	0	0	0	15	0	15	0	0	**70**	0	0	0	0	0	0	0
M = Drum	0	0	0	0	0	0	5	0	10	0	0	0	**85**	0	0	0	0	0	0
N = Trombone	0	0	0	0	0	0	0	0	0	0	0	0	0	**100**	0	0	0	0	0
O = Trumpet	5	0	5	0	0	0	0	0	0	0	0	0	0	5	**85**	0	0	0	0
P = Tuba	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	**95**	0	0	0
Q = Tympani	0	0	0	0	0	0	0	0	10	0	0	5	0	0	0	0	**85**	0	0
R = Viola	0	0	0	0	0	0	5	0	0	0	0	5	0	0	0	0	0	**90**	0
S = Violin	0	0	5	5	0	0	0	0	0	0	0	0	0	0	0	5	0	0	**85**

Table 6. Confusion matrix using the proposed features for instrument family classification using (all numbers are in %).

	String	Woodwind	Brass	Percussion
String	**88.54**	0	2.55	8.92
Woodwind	11.39	**73.42**	15.19	0
Brass	6	2	**91**	1
Percussion	0	8.4	0	**91.6**

reported for the woodwind family. The highest classification accuracy of 91.6% has been reported for the percussion family. An improvement of 8.52% has been seen in family classification because of additional information provided by HOS-based features.

4. Conclusions

In this paper, a new feature extraction technique has been proposed using hybridisation of Mel Frequency Cepstral Coefficients (MFCC) and HOS-based features for musical instrument classification. MFCC represents psychoacoustic properties of the human auditory system but does not provide instrument specific information. Higher Order Spectra (HOS)-based features have been used to provide instrument specific information such as non-linearity and non-Gaussianity of the instruments. These features have been derived from bispectrum and bicoherence values. Through experimentation it has been observed that when MFCC and HOS-based features are hybridised the classification accuracy of the instruments is improved significantly. An improvement of 6.39% for individual instruments and 8.52% for instrument families have been obtained because of instrument specific information provided by HOS-based features and auditory information provided by MFCC feature subset. Counter propagation neural network has been used as a classifier as it provides faster training as compared to other neural networks.

References

1. Agostini G., Longari M., Pollastri E. (2001), *Content-Based Classification of Musical Instrument Timbres*, International Workshop on Content Based Multimedia Indexing.

2. Agostini G., Longari M., Poolastri E. (2003), *Musical instrument timbre classification with spectral features*, EURASIP J. Appl. Signal Process., **1**, 5–14.

3. Ajmera P.K., Nehe N.S., Jadhav D.V., Holambe R.S. (2012), *Robust feature extraction from spectrum estimated using Bispectrum for speaker recognition*, Int. Journal of Speech Technology, **15**, 433–440.

4. Bhalke D.G., Rama Rao C.B., Bormane D.S. (2014), *Musical Instrument Classification using Higher Order Spectra*, International Conference on Signal Processing and Integrated Networks (SPIN-2014), 20–21 Feb., 2014.
5. Bhalke D.G., Rama Rao C.B., Bormane D.S. (2015). *Automatic Musical Instrument classification using Fractional Fourier Transform based-MFCC Features and Counter Propagation Neural Network*, Journal of Intelligent Information System, Springer publication, DOI: 10.1007/s10844-015-0360-9.
6. Bordoloi S., Sharmah U., Hazarika S.M. (2012), *Classification of Motor imagery based on Hybrid features of Bispectrum of EEG*, IEEE International Conference on Communications, Devices and Intelligent Systems (CODIS), pp. 123–116.
7. Byun H., Lee S.W. (2002), *Applications of support vector machines for pattern recognition*, [in:] Proc. of the International Workshop on Pattern Recognition with Support Vector Machine, pp. 213–236.
8. Choudhury M.A.S.S., Shah S.L., Thornhill N.F. (2002), *Detection and diagnosis of System Nonlinearities using higher order statistics*, 15th Triennial World Congress, Barcelona, Spain, pp. 1–6.
9. Deng J.D., Simmermacher C., Cranefield S. (2008), *A study on feature analysis for Musical Instrument Classification*, IEEE Transaction on Systems, Man and Cybernetics, **38**, 2, 429–438.
10. Dubnov S., Tishby N. (1994), *Spectral Estimation using Higher Order Statistics*, Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 1994.
11. Dubnov S., Tishby N. (1997), *Analysis of sound textures in musical and machine sounds by means of higher order statistical features*, International Conference on Acoustics, Speech, and Signal Processing, **5**, 3845–3848.
12. Dubnov S., Tishby N. (1998), *Testing for Gaussianity and Non Linearity in the sustained portion of musical sounds*, Recherches et Applications en Informatique Musicale, M. Chemillier, F. Pachet [Eds.], Editions HERMES, pp. 212–224.
13. Dubnov S., Rodet X. (2003), *Investigation of phase coupling phenomena in sustained portion of musical instruments sound*, J. Acoust. Soc. Am., **113**, 1, 348–359.
14. Eronen A. (2001), *Comparison of features for Musical instrument recognition*, [in:] Proceeding of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 19–22.
15. Eronen A., Klapuri A. (2000), *Musical Instrument Recognition using cepstral coefficients and temporal features*, [in:] Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, **2**, 753–756.
16. Essid S., Richard G., David B. (2006), *Hierarchical Classification of Musical Instruments on Solo Recordings*, [in:] Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, **5**, 14–19.
17. Goppert J., Rosenstiel W. (1993), *Self-organizing maps vs. back-propagation: An experimental study*, Proc. of Work. Design Methodol. Microelectron. Signal Process., pp. 153–162.
18. Goshvarpour A., Rahati S., Saadatian V. (2012), *Bispectrum Estimation of Electroencephalogram Signal During Meditation*, Iran J. Psychiatry Behav. Sci., **6**, 2.
19. Kaminskyj I., Czaszejko T. (2005), *Automatic Recognition of Isolated Monophonic Musical Instrument Sounds using kNNC*, Journal of Intelligent Information Systems, **24**, 2, 199–221.
20. Kostek B. (2004a), *Musical instrument classification and duet analysis employing music information retrieval techniques*, Proc. IEEE, **92**, 4, 712–729.
21. Kostek B. (2004b), *Application of soft computing to automatic music information retrieval*, Journal of American Society for Information Science and Technology, **55**, 12, 1108–1116.
22. Kostek B. (2007), *Applying computational intelligence to musical acoustics*, Archives of Acoustics, **32**, 3, 617–629.
23. Kostek B., Kania L. (2008), *Music information analysis and retrieval techniques*, Archives of Acoustics, **33**, 4, 483–496.
24. Kostek B., Czyzewski A. (2001), *Representing musical instrument sounds for their automatic classification*, Journal of Audio Engineering Society (JAES), **49**, 9, 768–785.
25. Kostek B., Krolikowski R. (1997), *Application of artificial neural networks to the recognition of musical sounds*, Archives of Acoustics, **22**, 1, 27–50.
26. Kostek B., Wieczorkowska A. (1997), *Parametric representation of musical sounds*, Archives of Acoustics, **22**, 1, 3–26.
27. Kuzmanovski I., Novič M. (2008), *Counter-propagation neural networks in Matlab*, Chemometrics and Intelligent Laboratory Systems, **90**, 84–91.
28. Li S., Liu Y. (2010), *Feature Extraction of Lung Sounds Based on Bispectrum Analysis*, Third International Symposium on Information Processing, pp. 393–397.
29. Liu R., Zolzer U., Guulemard M. (2010), *Excitation signature extraction for pitched musical instrument timbre analysis using Higher Order Statistics*, 2010 IEEE International Conference on Multimedia and Expo (ICME), 19–23 July 2010, 10.1109/ICME.2010.5582571.
30. Loughran R., Walker J., O'Neill M., O'Farrell M. (2008), *The Use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification*, International Computer Music Conference.
31. Martin K.D., Kim Y.E. (1998), *Musical Instrument recognition: A pattern recognition approach*, The Journal of Acoustical Society of America, **109**, 1768–1768, DOI: http://dx.doi.org/10.1121/1.424083.
32. Opolko F., Wapnick J. (1987), *MUMS – McGill University master samples* (in compact discs), McGill University, Montreal, Canada.