Multimedia Tools and Applications (2023) 82:33–58
https://doi.org/10.1007/s11042-022-13267-3

Empirical mode decomposition based statistical features for discrimination of speech and low frequency music signal

Arvind Kumar[1] · Mahesh Chandra[2]

Received: 2 July 2019 / Revised: 13 January 2021 / Accepted: 19 May 2022 /
Published online: 3 June 2022

Abstract

This work aims to investigate the significance of different Empirical Mode Decomposition (EMD) based statistical features for discrimination of speech and low frequency music signal (guitar signals) which mostly lie in the frequency range of 80–1200 Hz. Each of the speech/guitar audio samples is decomposed into 10 Intrinsic Function Mode (IMFs). These IMFs are further analyzed for discriminatory evidence using statistical features like Mean, Absolute Mean, Kurtosis, Variance and Skewness. These features are then fed to different classifiers and their performances were tabulated for varying tuning parameters of the classifiers. Initial experiments were conducted on isolated features to shortlist features with best discriminatory evidence. These shortlisted features were then used in different combinations and their performances were reported. An improvement of 19.13% is observed for hybrid features over isolated features. Speech samples were obtained from Scheirer and Slaney database and Guitar samples were generated from a continuous guitar monologue uploaded on YouTube. Feature selection technique using Fisher Method and F-ratio were also implemented and best feature vectors were reported for both the algorithm. Best overall accuracy of 82.16% is reported for Hybrid features with Radial Basis Function (RBF) kernel of SVM classifier when trained with top 38 feature vectors obtained using F-Ratio Method. Different experiments verified Absolute Mean and Variance as best performing features for our task.

Keywords Speech/music discrimination · EMD · IMF · Statistical features · Fisher method · F-ratio

✉ Arvind Kumar
arvind9835@gmail.com

Mahesh Chandra
shrotriya69@rediffmail.com

[1] Department of ECE, Birla Institute of Technology, Ranchi, India

[2] Department of ECE, Reva University, Bengaluru, India

1 Introduction

Audio segmentation and classification has been an important area of interest for many applications where initially the target could be to categorize the incoming signal into one of music, speech or silence. This could further be analyzed for various subdivisions, for example, into speaker identification or instrument identification or genre detection. This work focuses on classifying speech and guitar signal using Empirical Mode Decomposition and evaluates its performance for different classifiers. Although, many works have been carried out in the past for speech vs. music classification for broadcast news data but still classification of speech and guitar signal is an unexplored area. Guitar signals are example of low frequency music signal and broadly lie in the range of 80–1200 Hz. These kinds of signal are usually encountered during live play performance or beat poetry where an actor speaks between guitar monologues. Since both guitar and speech signal are non-stationary in nature sharing common spectrum in frequency domain, their classification is a complicated task.

1.1 Related works

Earlier work in the field of speech vs. music classification was conducted by Saunders [24] who used energy contour and zero-crossing (ZC) rate based statistical features for separation of speech and music. Accuracy up to 98% was reported when probability measures on signal energy was used over skewness of ZC rate distribution. Zhang and Kuo [35] proposed heuristic rule-based method for audio segmentation and classification in song, speech, environment noise and silence. Features like fundamental frequency, average ZC rate and spectral peaks tracks were used with an accuracy of more than 90% for audio classification and 95% for audio segmentation. Scheirer and Slaney [25] explored Power Spectral Density (PSD) in their work and proposed various features like spectral flux, spectral roll-off and spectral centroid for the task of speech and music discrimination. They tested these features with different classifiers and found an accuracy of 98.2% for 2.4 s segments. Alexandre et al. [1] used spectral based features along with Mel Frequency Cepstral Coefficients (MFCCs) and high zero crossing rate ratio with Fisher Linear Discriminant classifier and k-Nearest Neighbour. Another method proposed probability-based features using Hidden Markov Model (HMM) [31]. Gaussian likelihood ratio test was used for classification.

This work explores the use of Empirical Mode Decomposition for analysis of non-stationary signals to extract features to obtain discriminatory evidence between speech and guitar signals. The Empirical Mode Decomposition and the Hilbert spectrum had been extensively explored in the past for nonlinear and non-Stationary time series analysis [11]. EMD has been explored in earlier works for Speech/Music discrimination and promising results have been observed [13]. EMD has also been explored for detecting situational interest amongst students during learning [2]. Different frequency scales present in the signals are extracted using EMD which acts as a dyadic filter for the incoming signals [6, 32]. These extracted scales are also known as Intrinsic Mode Functions (IMFs). These IMFs have been explored for different applications like speech analysis, climate analysis, biomedical application, etc. because of the information embedded in them and promising results have been achieved [4, 9, 10]. Cepstral coefficient has been explored in [17, 26] for classification and segmentation of speech and music signal using Gaussian Mixture Model and SVM. Speech specific features for classification task have been proposed in [12]. A significant improvement is seen on combining speech specific features with existing features. Convolution Neural

Network using audio spectrogram has been proposed in [20]. Speech and music classification using IIR-CQT spectrogram based statistical descriptors and extreme learning machine has been proposed in [3]. A fast and efficient technique for segmentation and classification of speech and music signal using amplitude and Zero Crossing Rate (ZCR) is explored in [18]. Use of modified SVM for Speech/Music discrimination for Selectable Mode Vocoder (SMV) framework is explored in [16]. New feature vectors based on sinusoidal model for classification of speech and music signal is explored using SVM and GMM classifier in [28]. In [22], fundamental frequency is estimated for classification task. An audio-driven algorithm for the detection of speech and music events in multimedia content is introduced in [29]. D. Bykhovsky et al. improved robust voiced-unvoiced decision in presence of environmental noise using generalized likelihood ratio test (GLRT) [5]. Automatic threshold evaluation techniques were proposed in this work adapting both Constant false alarm rate (CFAR) and Bayes criterion thresholds.

1.2 Motivation

The aim of the study is to investigate and understand the efficacy of EMD based statistical features in classifying a speech and low frequency music signal sharing a common spectra range for different tuning parameters of state-of-the-art classifiers. A speech signal is governed by source-filter model which is almost similar across all speakers to produce speech. On the other hand, guitar signals are produced by vibration of strings in controlled manner. IMFs for both music and speech signals are expected to contain the information of the source. While, the IMFs for speech signals are expected to contain information of glottal activity in them, the IMFs for guitar signals should reflect the characteristics of vibrating strings. This fundamental difference in the production should result in different patterns of IMFs. The aim of this study was to extract features to exploit these differences and use it for classification of speech and guitar signals. The extracted features are tested with four different classifiers and results are compared.

Major objectives:

- To investigate and understand the efficacy of EMD based statistical features in classifying a speech and low frequency music signal sharing a common spectra range.
- To study the variation of the performance of the models on changing different tuning parameters of state-of-the-art classifiers.
- To analyze and sort best performing statistical features based on experimental results and inference.
- To study the improvement in performance of the models on passing different combination of best performing isolated features.
- To study the impact of feature selection technique on raw data and verify whether the manual interpretation of the best performing hybrid features proposed using experiments matches the results obtained from two feature selection techniques.

The rest of the paper is organized as follows: Section 2 briefs the process of EMD decomposition. Section 3 describes the preparation of database used for the experiment and presents analysis of extracted IMFs for speech and guitar signal. Section 4 describes the feature extractions process. Section 5 discusses the classifiers used in this work. Experimental results and observations are discussed in Section 6. Section 7 studies the effect of feature selection on performance of classifiers. Section 8 draws a comparative analysis of the present work with past works and Section 9 presents the conclusion and future scope of this work.

2 Empirical mode decomposition for audio signals

EMD has found its application in many real time analyses to extract AM-FM components of any complex signals breaking them into many IMFs [13]. This method decomposes any real time signals without any parametric optimization or without using a priori information. Table 1 describe the process of EMD in brief [13].

An IMF represents a single frequency scale satisfying the condition of having equal number of zero crossing and number of extrema or differs by at most one. Because of this rigid criterion, researcher had proposed several sifting criteria [30]. For this work, decomposition of the signals is limited to 10 IMFs to preserves the dyadic nature of EMD. Limiting the number of IMFs to 10 prevents unnecessary processing of latter IMFs which contain mostly low frequency trends for both speech and guitar signals.

3 Database and analysis of IMFs

Earlier work in music vs. speech discrimination mostly used Scheirer-Slaney database [25]. Since this work was focused on performance evaluation of EMD signal for speech and low frequency music signal classification; the latter couldn't be used in its original format. However, this work uses the speech samples contained in the data set leaving behind the music samples. Each of the speech samples was down sampled to 8 KHz from 22.05 KHz.

Table 1 Steps to evaluate EMD

Sl. No.	Description	Equations
1	For an original signal x (t)=r_o (t), (1) is used to obtain IMF h_k (t).	$r_{k-1}(t)=h_k(t)+r_k(t)$ (1) Where r_{k-1} (t) and r_k (t) are given residue obtained by *sifting process* to obtain IMF h_k(t).
2	Ideally, decomposition of a given signal into IMFs is stopped when residue takes the form of trend. A trend is a phenomenon in which the number of extrema in r_k(t) is 2 or less. But, practically, decomposition steps stop when the number of extracted IMFs reaches the user defined value 'M' where M is the number of required IMFs the audio signal is decomposed into.	$s(t) = r_m(t) + \sum_{k=1}^{M} h_k(t)$ (2) Where r_m is the residual signal and h_k(t) are M IMFs. In our study, we chose M=10 resulting in 10 IMFs
***shifting process* for a signal x(t)=r_o(t), whose IMFs needs to be extracted**		
1	For a residual signal r_{k-1}(t), minima and maxima envelop is constructed using cubic spline interpolation. Mean of these minima and maxima envelope gives the mean envelope m_{k-1}(t).	
2	Reducing this mean envelope from the original signal r_o(t), gives the new residue $r_k(t)= r_{k-1}(t)-m_{k-1}(t)$ (3)	
3	Above two steps (i) and (ii), are repeated till the sifting criterion is fulfilled or the IMF has reached a user defined value 'm' for decomposition.	

Guitar sound samples were downloaded from YouTube [34]. A continuous guitar wave file playing chords for 1 hour was downloaded and down sampled to 8 KHz. This wav file was broken into 80 samples each of 15 seconds to match the number of speech files in Scheirer-Slaney database. 60 of these 80 files from both speech and guitar databases were used for training and rest 20 files were used for testing. Spectral-Spread of these wav files was observed. Figure 1 shows the FFT of a sample of speech and guitar wav file. Most of the peaks in the FFT of speech signal range from 0 to 4000 Hz whereas for guitar signals, the peaks spanned mostly across 0–1500 Hz with small peaks occurring around 2600 Hz.

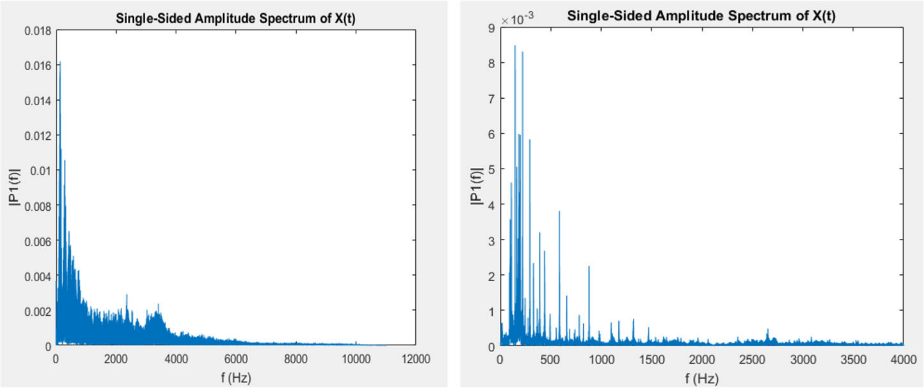
All the simulations were carried on with MATLAB R2015b running on Intel® Core™ i7–6700 64-bit processor with 8 GB RAM.

3.1 Analysis of extracted IMF

This section discusses the IMF’s extracted from speech and guitar signal. Figure 2 shows the first 7 IMFs extracted from speech and guitar audio samples respectively.

3.1.1 Analysis of IMFs of speech signal

The complex speech signal is decomposed into seven different IMF’s in decreasing order of frequency components. Works in the past have done AM-FM analysis of speech signals where the attempt is to represent a speech signal in terms of AM-FM components [19]. The AM-FM nature of the speech signal is clearly visible in first 3 IMFs in Fig. 5a. Sinusoidal waveforms are also seen spread across different IMFs especially in IMF 5, 6 and 7. These sinusoidal reflect the voiced speech segments and have been used by researcher to find glottal activity [27]. However, the task becomes difficult because of mode-mixing where a frequency scale is distributed among different IMF’s and different frequency scales are merged in one IMF [2, 8]. This problem is solved by advanced version of EMD like Ensemble-Emperical Mode Decomposition (EEMD) [33] and its variants. EEMD works by adding small amplitude white noise to the original data and take the ensemble mean of the IMFs extracted from this noisy data. Over much iteration, this added white noise is averaged out leaving behind the original components of the signal simultaneously separating the modes into its proper IMFs.



(a) (b)

Fig. 1 Single Sided Amplitude Spectrum of (**a**) Speech Signal (**b**) Guitar Signal

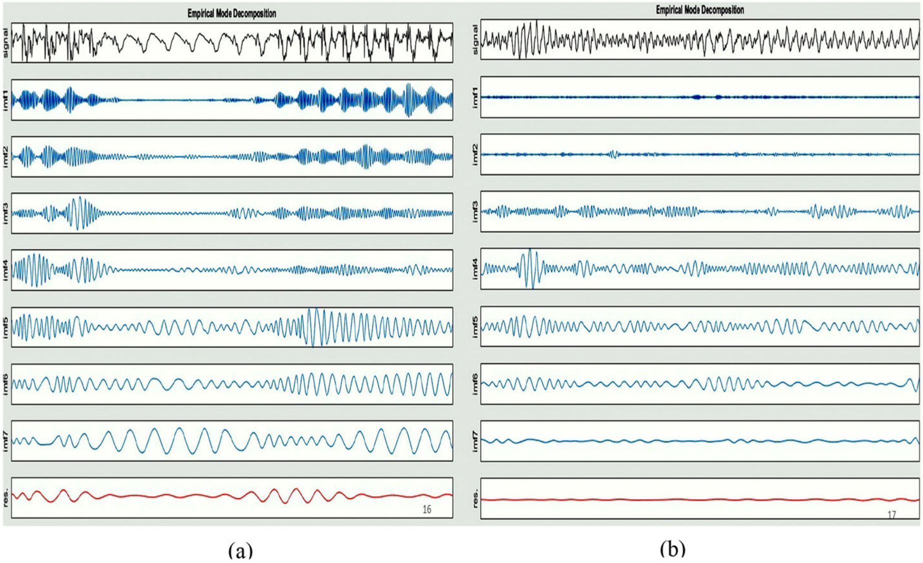


Fig. 2 IMF 1–7 from EMD decomposition of (**a**) Speech Signal (**b**) Guitar Signal

3.1.2 Analysis of IMFs of guitar signal

Unlike, speech samples which are governed by source filter theory, the source of music sound is completely different and hence the difference in the IMF's of the two was expected. Guitar wave files had musical chords recorded in them which are played by vibrating three or more notes simultaneously. The presences of different frequency components are clearly visible in first 6 IMF's in Fig. 2b. A clear difference is seen in the first two IMF's of speech and guitar signal. While the latter reflected AM-FM nature of speech, very few high frequency components were observed in the IMFs extracted from guitar signal. The IMFs from guitar were more sinusoidal in nature and mode-mixing can be clearly seen in IMF 3 to 6. A low frequency trend like waveform is observed for IMF 7 for guitar signal unlike for speech sample which have oscillation for residue too.

4 Feature extraction

The main objective of this work was to perform statistical data analysis on the IMF's generated from Empirical Mode Decomposition of speech/guitar signal and observe the discriminatory characteristics in them. These can be used for machine learning to solve the classification problem. In the past, such statistical features have shown significance performance in different classification problems, especially in the field of EEG, ECG, speech, and music signal processing [2, 13, 15]. Statistical data analysis aims to quantify the data by applying various statistical operations for efficient use of data by classification algorithms. In this work, 5 different statistical operations are used. These are mean, absolute mean, variance, skewness and kurtosis.

Figure 3 illustrates the training and testing of models using audio samples each of 15 seconds. Training samples of speech/guitar signals were pre-processed and down-sampled to 8 KHz. The down-sampled signals are fed to Emperical Mode Decomposition (EMD) algorithm to generate ten IMFs. These IMFs are chopped into non-overlapping frame of one second resulting into 15 frames. These smaller frames are fed to feature extraction block where five different statistical features were evaluated. Hence, every training and testing speech/guitar sample signal was represented by a matrix of size 10 × 15 i.e., 15 frames of one second for 10 IMFs. These features are then individually normalized and are fed to different classifiers, along with target labels for training the models. Once the model is successfully trained, features from test data are fed to the trained model to label them into either of speech/guitar class. Table 2 tabulates the features used.

5 Classifiers

This work is focused on binary classification task. Previous work in speech/music discrimination has extensively used Support Vector Machine (SVM) and k-Nearest Neighbour (k-NN) and has found satisfactory results [15, 25]. Hence, it was a motivation to test the performance of these two classifiers for this task. Along with these, some experiments were also run on Naïve Bayes and Artificial Neural Network classifiers. A comparative analysis is tabulated in the next sections.

5.1 Support Vector Machine

It is a discriminative classifier which separates two different classes by a hyperplane for a given vector weight w and bias b. Earlier work have explored the performance of SVM for speech/music classification [14]. The distance between the closest data points and the hyperplane is called margin of separation. These points which are closest to the hyperplane are called support vectors. The algorithm tries to find a hyperplane which maximizes the margin of

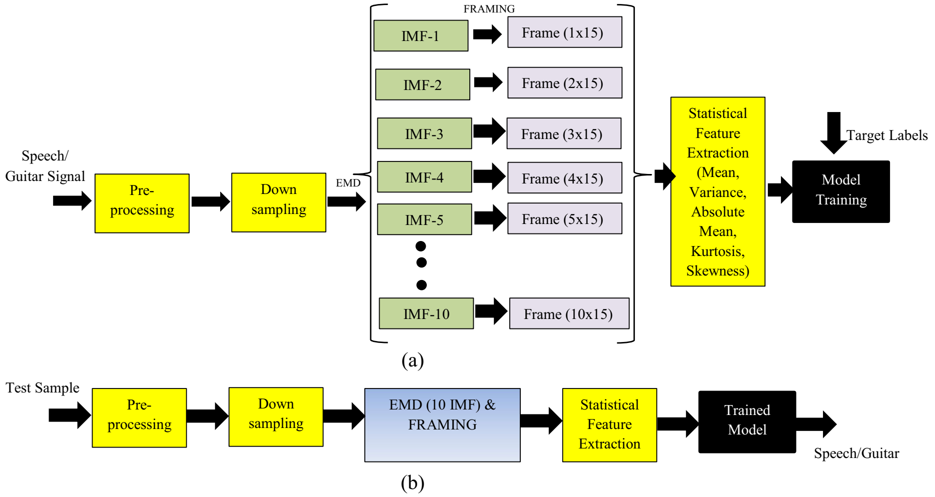


Fig. 3 Speech & Guitar Signal Classification (**a**) Training (**b**) Testing

Table 2 Description of features

Sl. No.	Feature	Description	Formula
1	**Mean**	Mean of a sample values are the sum of the sampled values divided by the number of samples. It is denoted by $\overline{x}$.	Mean of x_1, x_2, ……..x_n is given by $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ (4)
2	**Absolute Mean**	Absolute Mean of a sample values is the sum of the absolute value of sampled values divided by the number of samples. It is dented by $\overline{\lvert x \rvert}$.	Absolute Mean of x_1, x_2, ……..x_n is given by $\overline{\lvert x \rvert} = \frac{1}{n}\sum_{k=1}^{n} \lvert x_k \rvert$ (5)
3	**Variance**	Variance of a sample values measures the spread of the sample values from its mean. It is the expectation of squared deviation of a random variable from its mean.	Variance of random variable X is given be $V = E\left[(X - \overline{x})^2\right]$ (6) Where, E (t) is the expected value of t and $\overline{x}$ is the mean. For a random variable vector, A made up of N scalar observations, the variance V is defined as $V = \frac{1}{n-1}\sum_{k=1}^{n} \lvert A_k - \overline{x} \rvert^2$ (7) Where, $\overline{x}$ is the mean of the A.
4	**Skewness**	Skewness measures the asymmetry nature of the data around the mean. For data which are spread out more to the left of the mean than to right, skewness is negative and if the data are spread more to the right of the mean than to left, the skewness is positive. A perfectly symmetrical distribution gives zero skewness.	Skewness of random variable X is given by $S = \frac{E\left[(X - \overline{x})^3\right]}{\sigma^3}$ (8) Where, E (t) is the expected value of t, $\overline{x}$ is the mean and σ is the standard deviation. For a random variable vector, A made up of N scalar observations, the skewness S is defined as $S = \frac{\frac{1}{n}\sum_{k=1}^{n} \lvert A_k - \overline{x} \rvert^3}{\left(\sqrt{\frac{1}{n}\sum_{k=1}^{n} \lvert A_k - \overline{x} \rvert^2}\right)^3}$ (9) Where, $\overline{x}$ is the mean of the A.
	Kurtosis	Kurtosis measures the outlier-proneness of a distribution. Kurtosis of distributions with more outlier-proneness than a normal distribution has K value greater than 3 whereas kurtosis of distributions with less outlier-proneness than a normal distribution has K value less than 3.	Kurtosis of random variable X is given by $K = \frac{E\left[(X - \overline{x})^4\right]}{\sigma^4}$ (10) For a random variable vector, A made up of N scalar observations, the Kurtosis K is defined as $K = \frac{\frac{1}{n}\sum_{k=1}^{n} \lvert A_k - \overline{x} \rvert^4}{\left(\sqrt{\frac{1}{n}\sum_{k=1}^{n} \lvert A_k - \overline{x} \rvert^2}\right)^4}$ (11) Where, σ is the standard deviation.

separation. In 2D, this hyperplane is a line which divides the plane in two different parts where each class lies on either side. For more complex data which are not linearly separable in two dimensions, SVM uses kernels to map the data in higher dimension. Three different kernels namely Linear, Radial Basis Function (Gaussian) and Polynomial, were experimented for this work and a comparative analysis was done. Eq. 12 represents the general equation of SVM. Eqs. 12–14 represents the different kernels used in this experiment. All the simulations were done with *fitcsvm* function in MATLAB. *Fitscsvm* uses Sequential Minimal Optimization

(SMO) as a solver for binary classification and optimally finds the width parameter Y and the cost parameter c. For a set of data with training vectors x_j and their categories y_j in some dimension d where $x \in R^d$ and $y_j = \pm 1$, the equation of hyperplane is

$$f(x) = x^{'}w + b = 0 \tag{12}$$

Where, w and b are weight vector and bias respectively.

5.1.1 Non-linear transformation using kernels

As discussed earlier, when simple hyperplane fails to classify some problems, variant of mathematical approaches are used which retains all the property of an SVM separating hyperplane. For a class of functions $G(x_1, x_2)$, a function φ *maps x to linear space S such that.*

$$G(x_1, x_2) = < \varphi(x_1), \varphi(x_2) > \tag{13}$$

The functions used are:

(i). Polynomials: For some positive integer *p*,

$$G(x_1, x_2) = \left(1 + x_1{'}x_2\right)^p \tag{14}$$

For this experiment, p = 2 has been used.

(ii). Radial basis function (Gaussian)

$$G(x_1, x_2) = \exp.(-\|x_1 - x_2)\|^2\Big) \tag{15}$$

5.2 K-Nearest Neighbours

K-Nearest Neighbours (k-NN) classifier uses data spread in multidimensional feature space each having class labels for training. It is not only easy to interpret; it takes very less computation time. ‘k’ in k-NN is a user defined constant and indicates the number of neighbors voting is made from. A test sample is assigned a class by assigning the label of the training samples which occurs maximum number of times among the training samples closest to the test point. Use of k-NN has been explored in [25]. This algorithm finds the distance of the test samples with k-nearest neighbors. Generally, Euclidean distance is a commonly used distance metric. However, in this work, performance of Chebychev and Mahalanobis distance metric is also observed. All the simulations were done with *fitcknn* function in MATLAB.

(i). Euclidean Distance

For a given set of points (x_1, y_1), (x_2, y_2), Euclidean distance is given by

$$d = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \tag{16}$$

(ii). Chebychev Distance

For a given set of points (x_1, y_1), (x_2, y_2), Euclidean distance is given by

$$d = max|(y_2 - y_1), (x_2 - x_1)| \tag{17}$$

(iii). Mahalanobis Distance

For a given vector of data x, Mahalanobis distance is given by

$$d^2 = (x - m)^T C^{-1} (x - m) \tag{18}$$

Where m is the mean of variables and C^{-1} is inverse covariance matrix. Mahalanobis Distance transforms the variable into uncorrelated variable and makes their variance equal to 1. It normalized the variation of spread among variables and find simple Euclidean distance.

5.3 Naïve Bayes Classifier

Naïve Bayes Classifier is based on application of Bayes Theorem. These are simple probabilistic classifiers based on assumptions that there is strong independence between features. Using Bayes Rule,

$$P(Y/X_{1,..........}, X_n) = \frac{P\left(\frac{X_{1,.........}, X_n}{Y}\right)P(Y)}{P(X_{1,........}, X_n)} \tag{19}$$

Where,

$P\left(\frac{X_{1,..........}, X_n}{Y}\right)$ = Likelihood Probablity, $P(Y/X_{1,............}, X_n)$ = Posterior Probability, P(Y) = Prior Probability,

X_1,…….., X_n = Set of feature vectors.

If all the features are independent,

$$P\left(\frac{X_{1,.........}, X_n}{Y}\right) = \prod_{i=1}^{n} P\left(\frac{X_i}{Y}\right) \tag{20}$$

This reduces computation complexity. Using these equations, a Naïve Bayes Probability model is generated which is combined with a decision rule. One such rule is to select the most probable outcome which is also known as maximum a-posteriori rule. Simulations were carried on with *fitcnb* function in MATLAB.

5.4 Artificial Neural Network

Artificial Neural Network is a framework for machine learning inspired by biological neural network. It has an input layer, hidden layers and an output layer. Features are fed to the framework from the input layer which is fed to well-connected hidden layers. These hidden layers finally dump the data into the output layer which is a softmax layer. Each of these nodes has a weight vector and a bias. Summation of the product of input vector and weight vector with bias is fed to an activation function which is generally a sigmoid function given by Eq.

21. The difference between target output and evaluated output is the error signal which is fed back into the system for weight update. Numbers of hidden layers are varied in this experiment to see its impact on the classification efficiency. MATLAB function *nnstart* is used to simulate the experiment. This tool uses scaled conjugate gradient back propagation method for training.

$$f(x) = 1/(1+e^{-x}) \tag{21}$$

Softmax function is given by standard exponential function on each of the variables divided by sum of the exponential function for each variable. This acts as a normalizing constant and sums the output variables to 1. Eq. 22 represents a softmax function

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \; for\ j = 1\ to\ K \tag{22}$$

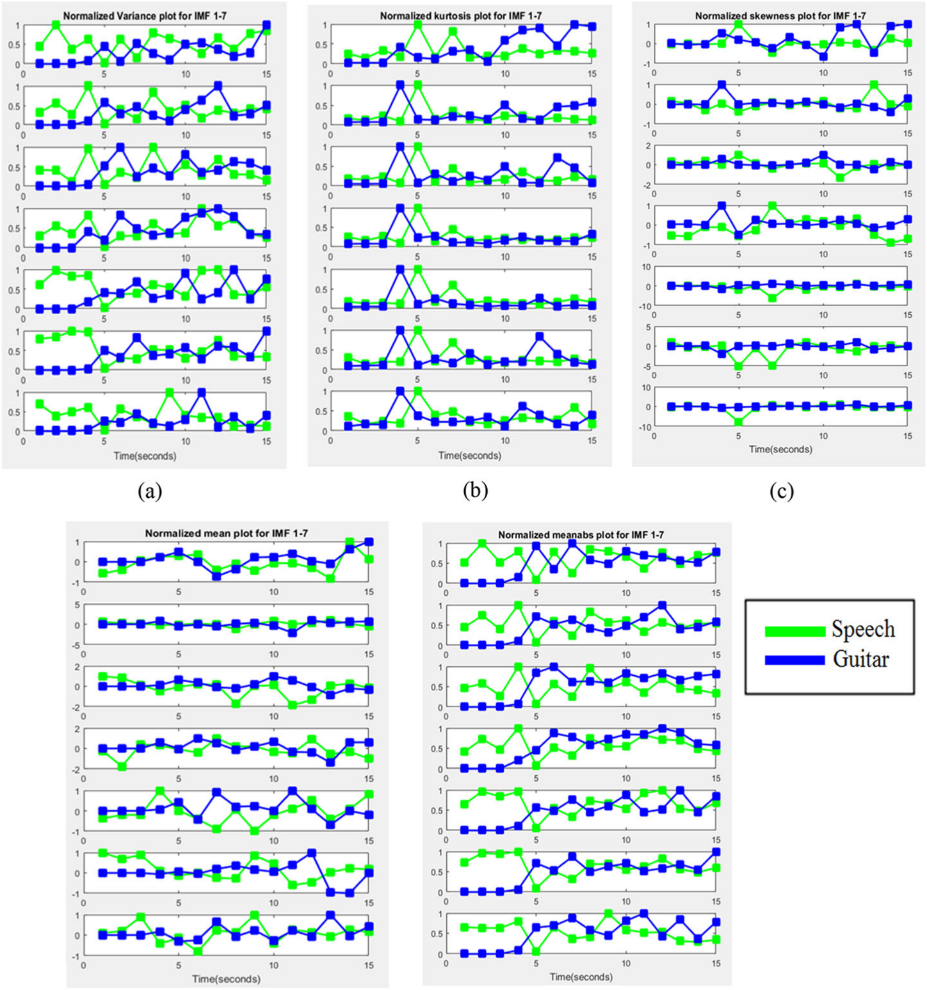
6 Simulation results & discussion

This section presents the simulation results. Initial experiments were run on isolated features to select features with best discriminatory evidence. This was followed by study on hybrid features.

6.1 Isolated features

Initially, all the five different features were analyzed independently for their discriminatory nature across first 7 IMFs. Figure 4 displays the line plot of the normalized value of all the five different feature set for a sample speech and guitar file computed over 15 seconds. Green line represents the feature set for speech signal while blue line represents the feature set for guitar signal. Figure 4a shows the variance plot of the IMF's. Variance of speech signal has comparatively higher value than that of guitar signal for most of the sample values in IMF 1 and 2. Variance plot for IMF 3 to 7 also showed good discriminatory evidence with very few samples merging on each other. On the other hand, plot of Kurtosis and Skewness showed a lot of correlation between feature set for speech and guitar signal, especially for IMF 4 to 7. Poor discriminatory evidence was also seen for IMF 1–4 for Mean feature. IMF 5–7 showed satisfactory results for the same. Good discriminatory evidence was seen for the plot of absolute mean for almost all the IMFs. The conclusions drawn from Fig. 4 were further validated from the scatter plot. Figures 5 and 6 shows the scatter plot of all the features computed for IMF 2 and 3 and also for IMF 5 and 6 respectively. It is clearly evident from Figs. 5 and 6 that absolute mean and variance shows good discriminatory evidence for classification of speech and guitar signal. Kurtosis feature shows better discriminatory evidence for IMF 5 and 6 than for IMF 2 and 3. Both Kurtosis and Skewness features were less spread in the plot with many data points merging over each other. Mean feature also shows good discriminatory evidence in both plots in Figs.5 and 6. These features were then fed into different classifiers and their performances were evaluated.

Classification accuracy was considered as the parameter for evaluation and is explained in Eq. 23. The numbers in the following tables represent the classification accuracies of the



(a) (b) (c)

(d) (e)

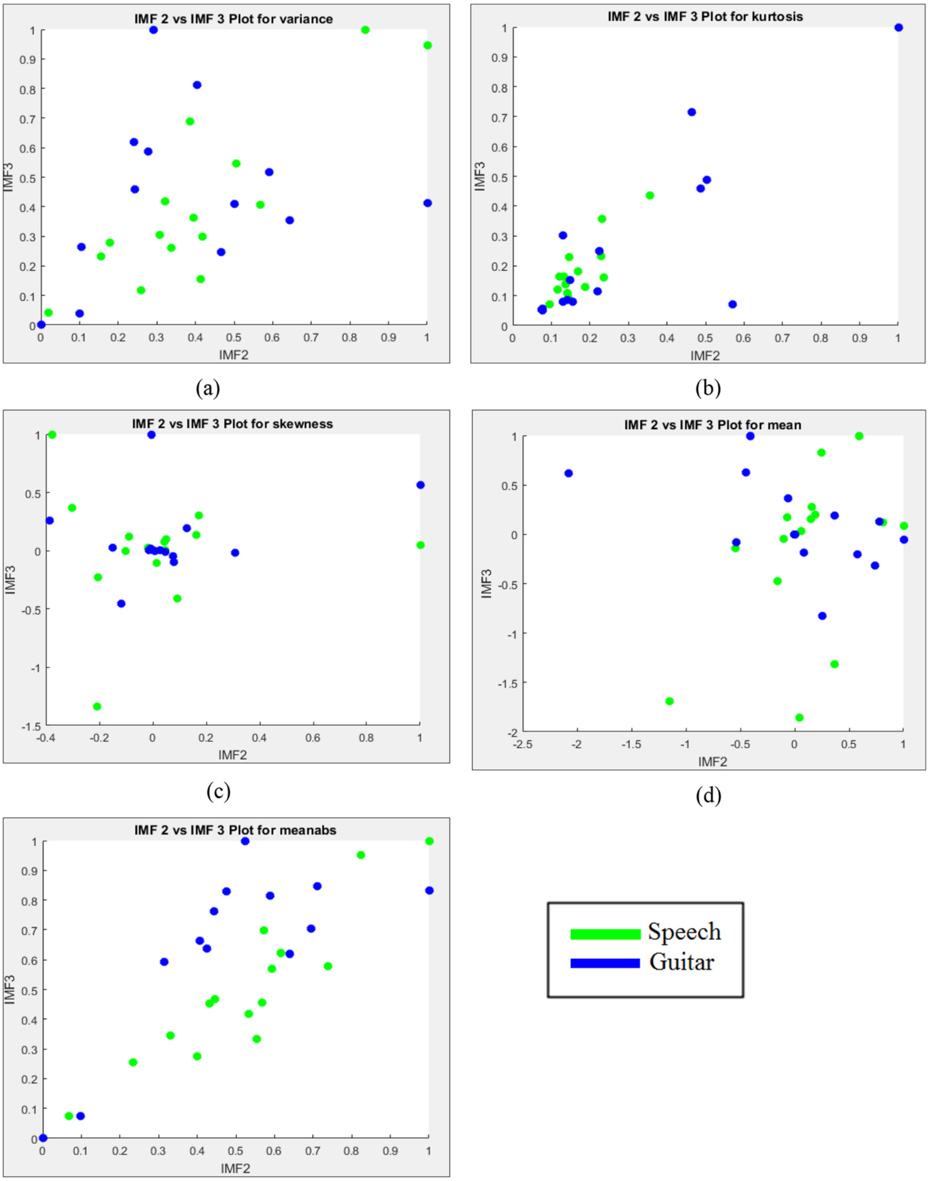
Fig. 4 Normalized line plot for (**a**) Variance (**b**) Kurtosis (**c**) Skewness (**d**) Mean (**e**) Absolute Mean

models across different features and classifiers parameters in percentage (%). Overall column represent the net classification accuracy of the model (in %) found by averaging the classification accuracies of individual classes (Speech and Guitar).

$$Classification\ accuracy(in\%) = (TP + TN)/(TP + TN + FP + FN)^{*}100 \qquad (23)$$

Where, TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

Table 3 displays the performance evaluation of SVM classifiers for five different features across three different kernels. Absolute Mean feature out-performed all other features with a classification accuracy of 68.83% when used with polynomial (order =2) kernel. Absolute mean and variance performed equally well for classification of both speech and guitar signal unlike skewness which performed better only for guitar signal. Performance of mean and kurtosis were satisfactory. Performance evaluation of SVM classifiers for different feature



(a)

(b)

(c)

(d)

(e)

Fig. 5 Scatter plot for (**a**) Variance (**b**) Kurtosis (**c**) Skewness (**d**) Mean (**e**) Absolute Mean on IMF 2 and 3

vectors with different kernels was also validated using Receiver Operator Characteristic (ROC) curve in Fig. 7. The parameter to judge the efficiency of a feature is Area Under the Curve (AUC). A perfect feature set will have AUC equal to 1. Curve closer to upper left corner will be comparatively better than other feature sets. Best results were observed for Radial Basis Function (RBF) kernels for kurtosis, absolute mean and variance feature and are reflected in their ROC plot in Fig. 7c.

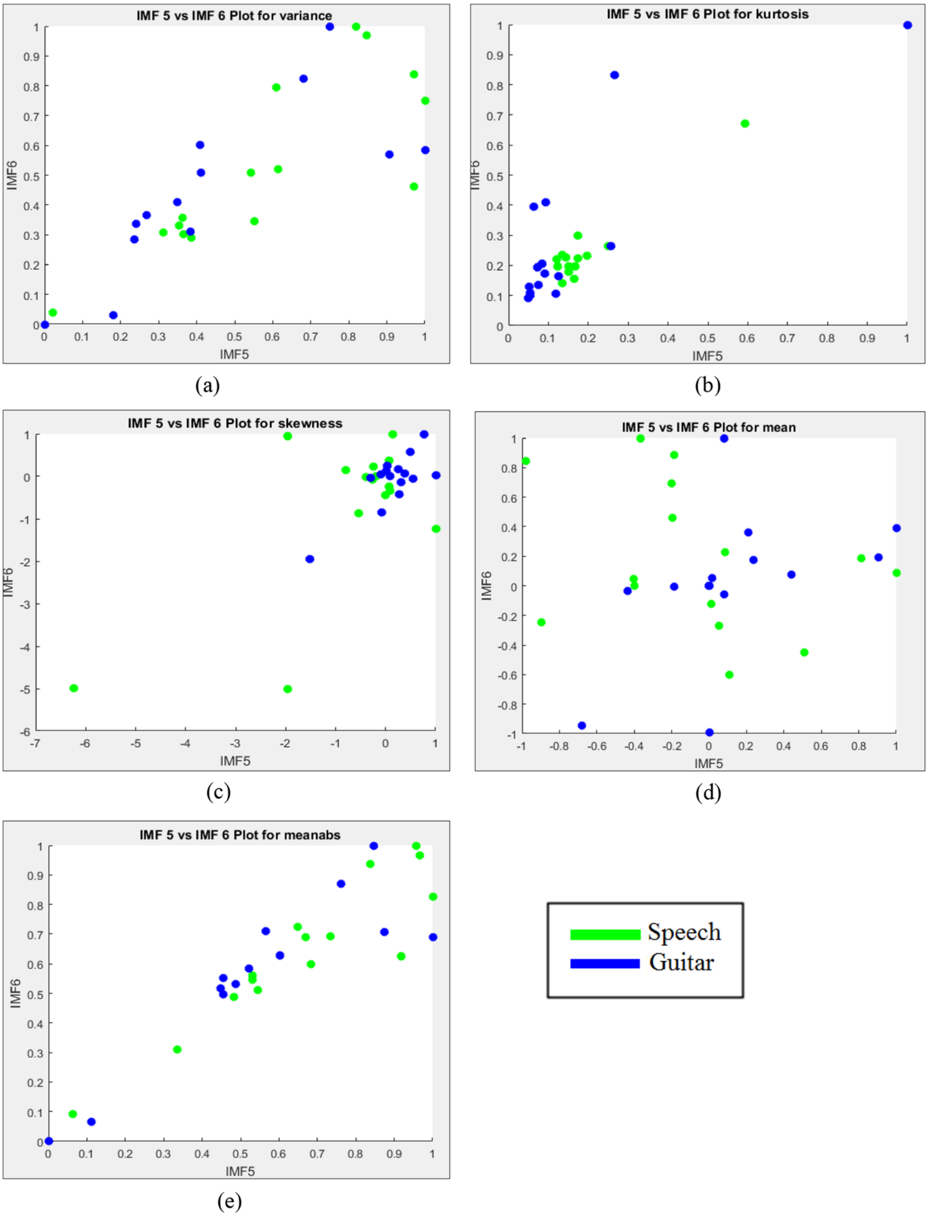


Fig. 6 Scatter plot for (**a**) Variance (**b**) Kurtosis (**c**) Skewness (**d**) Mean (**e**) Absolute Mean on IMF 5 and 6

Table 4 displays the performance evaluation of KNN classifiers for three different distances metric. Euclidean distance metric performed comparatively best amongst the three followed by Mahalanobis and Chebychev distance metric. For KNN, better classification accuracy was observed for guitar signal in all three scenarios. Absolute mean feature had the best results amongst the five features used with classification accuracy of 62.83%. Variance, skewness and kurtosis feature performed satisfactorily. Mean feature showed the least performance efficiency amongst all.

Table 3 Classification accuracy of SVM classifier (in %) for different Kernels

SVM	Linear			RBF			Polynomial (order 2)		
Feature	Speech	Guitar	Overall	Speech	Guitar	Overall	Speech	Guitar	Overall
Kurtosis	58.00	39.34	48.67	61.34	53.67	57.50	74.00	48.34	61.17
Variance	59.67	61.67	60.67	55.00	70.66	62.83	60.67	70.34	65.50
Skewness	13.34	85.00	49.17	57.00	44.67	50.83	25.34	82.00	53.67
Abs. Mean	53.00	71.34	**62.17**	63.00	73.67	**68.33**	59.67	78.00	**68.83**
Mean	43.00	76.00	59.50	50.67	42.67	46.67	41.67	66.67	54.17

Table 5 displays the performance evaluation for Naïve Bayes Classifier. Best results were seen for variance with overall classification accuracy of 55.99%. While kurtosis performed well for classification of speech signal, skewness and mean performed exceptionally well for classification of guitar signal with classification accuracy of 88.33% and 95.00%. Absolute mean stood as a second best feature with overall classification accuracy of 54.16%. Figure 8

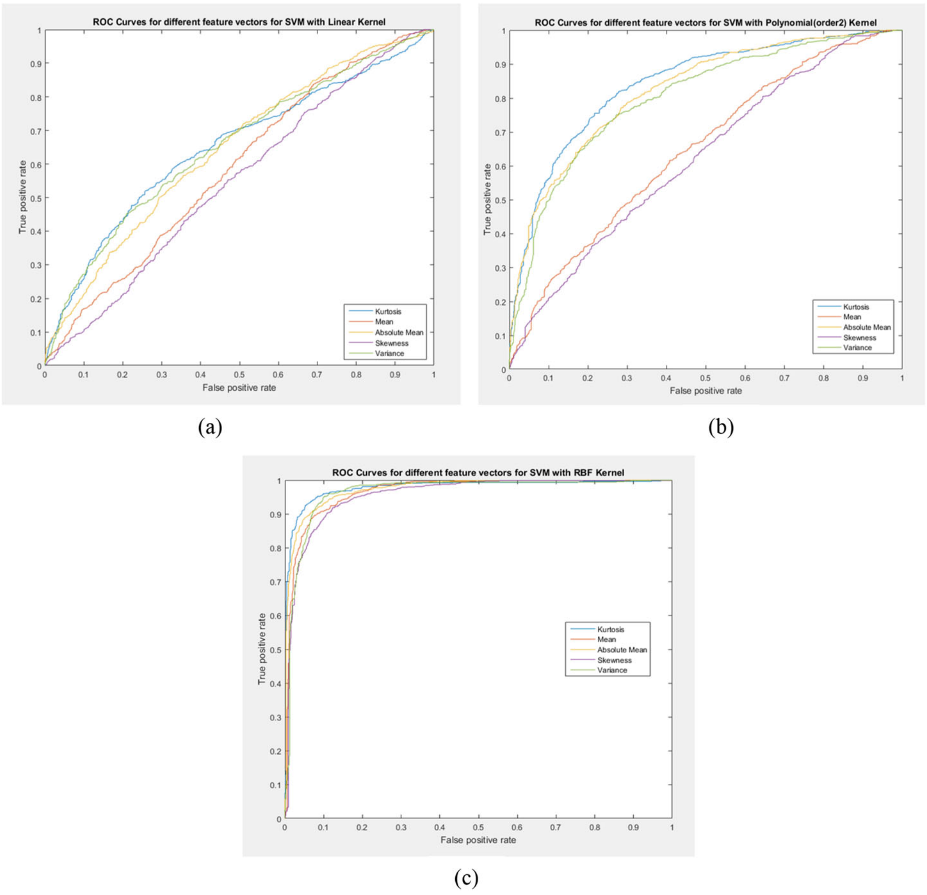


Fig. 7 ROC Curve for different feature vectors for SVM with (**a**) Linear (**b**) Polynomial (**c**) RBF Kernel

Table 4 Classification accuracy of KNN classifier in % for different distance metric

KNN	Euclidean			Chebychev			Mahalanobis		
Feature	Speech	Guitar	Overall	Speech	Guitar	Overall	Speech	Guitar	Overall
Kurtosis	48.66	61.67	55.16	46.34	61.67	54.00	45.66	65.33	55.49
Variance	35.00	79.00	57.00	34.00	76.67	55.33	34.33	77.34	55.83
Skewness	22.34	82.00	52.17	22.00	81.67	51.83	21.00	83.34	52.17
Abs. Mean	47.33	78.33	**62.83**	39.67	77.00	**58.33**	38.33	81.00	**59.66**
Mean	24.66	73.34	48.99	24.67	74.34	49.50	23.34	74.00	48.67

shows the ROC plot for different feature vectors with Naïve Bayes Classifier. Average performance was observed with poor AUC.

Table 6 display the performance evaluation of ANN for 3 different numbers of hidden layers. While performance for some of the feature vectors improved as N is increased from 5 to 10, performance for others saw a decline. Classification accuracy for Variance and Skewness improved as N is increased from 5 to 10. Skewness and Absolute Mean saw a dip in its performance from 52.83% and 65.33% to 48% and 58.84%. As N is further increased to 20, performance for all the features saw a decline except Kurtosis which saw an improvement of 6.64%. Best results were observed for Variance with an accuracy of 68% for N = 10.

Comparative results for all the four classifiers are presented in Table 7 and Fig. 9. Best results were observed for the combination of SVM and Absolute Mean with an accuracy of 68.83%. Amongst the classifiers, SVM performed best with an overall accuracy of 61.73%, followed by ANN (61.43%), KNN (55.49%) and Naïve Bayes (52.56%). Amongst the features, the best performing feature was Absolute Mean with an overall accuracy of 62.78% followed by Variance (61.62%), Kurtosis (56.45%) and Skewness (52.29%).

6.2 Hybrid features

From the above study on the use of isolated features with different classifiers for the task of guitar and speech signal classification, it was concluded that Absolute Mean and Variance stood as best two performing features. The experiments were continued with hybrid features concatenating two or more features and their classification accuracies were observed. To verify the discriminatory characteristics of best two performing feature i.e. Absolute Mean and Variance, a scatter plot is drawn. Figure 10 shows the scatter plot of Absolute Mean vs. Variance for two different IMFs. The discriminatory evidence is prominent in both. Comparative results for hybrid features are tabulated in Table 8. A sharp improvement is seen in performance of all classifiers when used with feature combination of Absolute Mean and Variance. Best results were observed for SVM (polynomial kernel) when the model is trained

Table 5 Classification accuracy of Naïve Bayes Classifier (in %)

Feature	Speech	Guitar	Overall
Kurtosis	69.66	25.66	47.66
Variance	61.66	50.33	**55.99**
Skewness	12.66	88.33	50.49
Abs. Mean	65.00	43.33	54.16
Mean	14.00	95.00	54.50

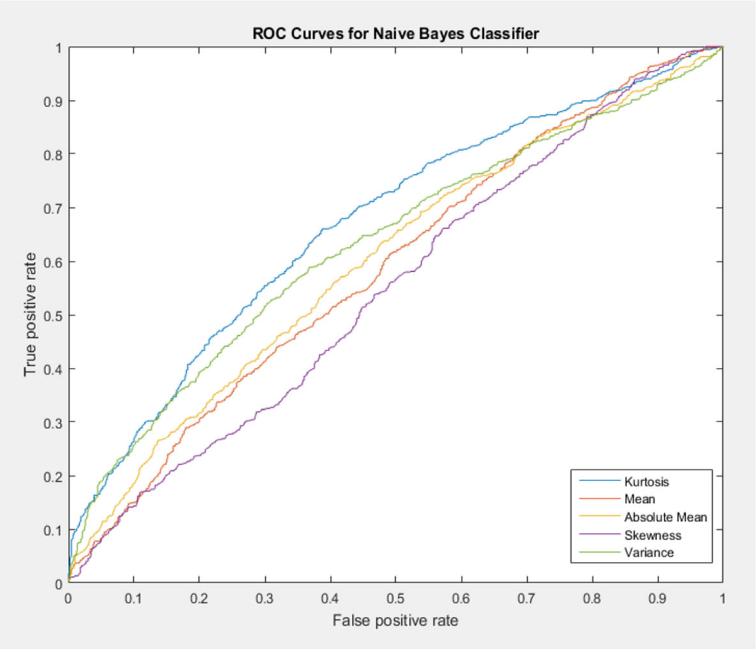


Fig. 8 ROC Curve for Naïve Bayes Classifier

with hybrid of Absolute Mean, Variance and Kurtosis Features (82.00%). Figure 11 displays the ROC plot of the same. AUC for RBF and Linear kernel indicates promising results.

6.3 Feature selection

Data reduction is the art of reducing the problem of high dimensionality to improve computational complexity and data acquisition cost by selecting most efficient features with maximum discriminatory evidence. In this study, two different techniques of feature selection are used.

6.3.1 Feature selection using fisher method

Fisher method assigns a score for a feature. This score is the ratio of interclass separation and intra-class variance [7]. Final feature selection occurs by segregating the *m* top ranked features,

Table 6 Classification accuracy of ANN(in %) for different no. of hidden layers

ANN	N=5			N=10			N=20		
Feature	Speech	Guitar	Overall	Speech	Guitar	Overall	Speech	Guitar	Overall
Kurtosis	63.00	35.33	49.16	61.34	54.00	57.67	74.34	48.67	61.50
Variance	62.66	63.67	63.16	62.00	74.00	**68.00**	61.00	72.67	**66.83**
Skewness	47.00	58.66	52.83	35.67	60.34	48.00	31.00	63.00	47.00
Abs. Mean	62.67	68.00	**65.33**	39.34	78.34	58.84	69.67	48.00	58.83
Mean	87.67	31.33	59.50	41.00	76.67	58.83	52.00	65.00	58.50

Table 7 Comparative performance of different classifiers (in %)

	SVM	KNN	NB	ANN
Kurtosis	61.17	55.49	47.66	61.50
Variance	65.50	57.00	55.99	68.00
Skewness	53.67	52.17	50.49	52.83
Abs. Mean	**68.83**	62.83	54.16	65.33
Mean	59.50	50.00	54.50	59.50

where *m* is a user defined constant ranging from one to total number of features. Fisher method was applied over the statistical raw features and its performances were evaluated for different values of *m*. Feature Selection Library (FSlib 2018) were used for simulating the results [21]. Table 9 tabulates the variation of efficiency of different SVM models with change in number of feature vector used. Highest efficiency of 80.33% is seen when RBF kernel is trained with 49 features. Figure 12 summarizes the ranking order amongst the feature vectors. Performance of varying number of hybrid feature vectors selected using Fisher Method with SVM (rbf) classifier can be seen in ROC curve in Fig. 12. As number of features used is increased from 10 to 50, AUC also increases, indicating improvement in results (Fig. 13).

Distributions of different feature vectors over rank 1 to 50 were studied and a histogram was plotted. Best performing features were Variance, Absolute Mean and Mean with maximum occurrence amongst top ranking vectors. Skewness showed maximum visibility across rank 21–30. Kurtosis was spread across the histogram with most occurrences across rank 41–50. This study confirms the order of discriminatory evidence amongst the statistical feature used. The discriminatory evidence amongst the feature vectors was further investigated using F-ratio.

6.3.2 Feature selection using F-ratio

F-ratio is a measure of variance of multi-class data [23]. It is the ratio of the variance of means between classes and the average variance within each class. Mathematically,

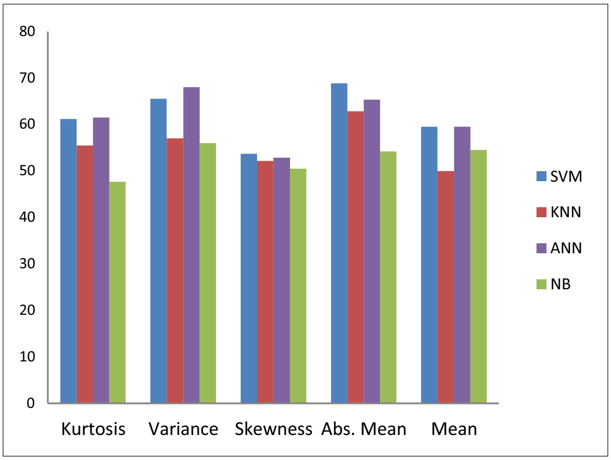


Fig. 9 Performance of different classifiers

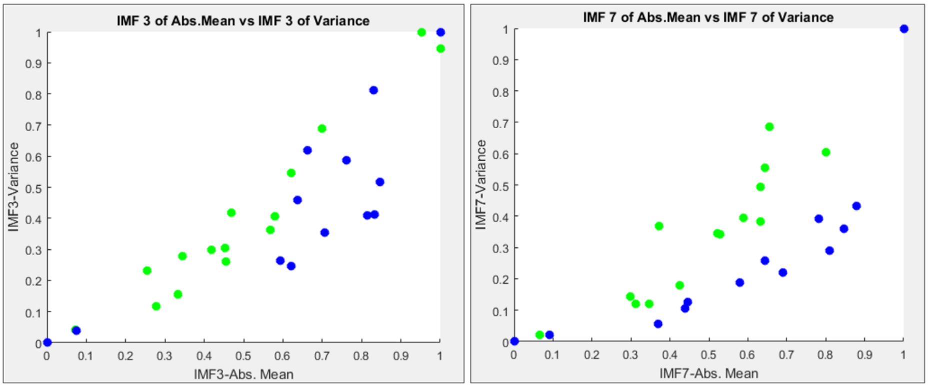


Fig. 10 Scatter plot for Absolute Mean and Variance for (**a**) IMF3 (**b**) IMF7

$$F\text{−}ratio = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(\mu_j - \overline{\mu}\right)^2}{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n_j}\sum_{i=1}^{n_j}\left(x_{ij} - \mu_j\right)^2} \tag{24}$$

Where, k is total number of class, μ_j is mean of jth class, $\overline{\mu}$ is total mean, n_j is total number of data points in a class and x_{ij} is ith data point in jth class. A higher F-ratio indicates more similarity within a class and more dissimilarity across the class. Following steps were performed for finding optimum features using F-ratio.

Step 1. F-ratio for all the 5 statistical feature vectors is evaluated using (24).
Step 2. Maximum of these F-ratio is found.
Step 3. Threshold value for F-ratio is given by Eq. 25, where k is varied from 1 to 100.

$$Thres = (F\text{−}ratio)/k \ \text{ where}, 0 < \text{k} < 100 \tag{25}$$

Step 4. Feature Vectors with F-ratio more than the set threshold value is selected for training and testing the model. Results are tabulated in Table 10.

To evaluate the performance of SVM model, *k* is varied from 1 to 100 changing the threshold value. As '*k*' is increased decreasing the threshold, more numbers of feature vectors were appended to the training matrix. The number of training vectors for varying k repeated

Table 8 Performance comparison for Hybrid Features (in %)

Feature	SVM (polynomial)	SVM (rbf)	SVM (linear)	KNN (Euclidean)	ANN (N=20)
All five features	77.66	78.83	72.00	60.50	78.83
Absolute Mean+Variance	79.67	74.16	72.16	**70.50**	78.83
Absolute Mean+Variance+Mean	74.80	76.67	**74.67**	61.00	76.50
Absolute Mean+Variance+kurtosis	**82.00**	**79.33**	71.67	61.16	**80.66**

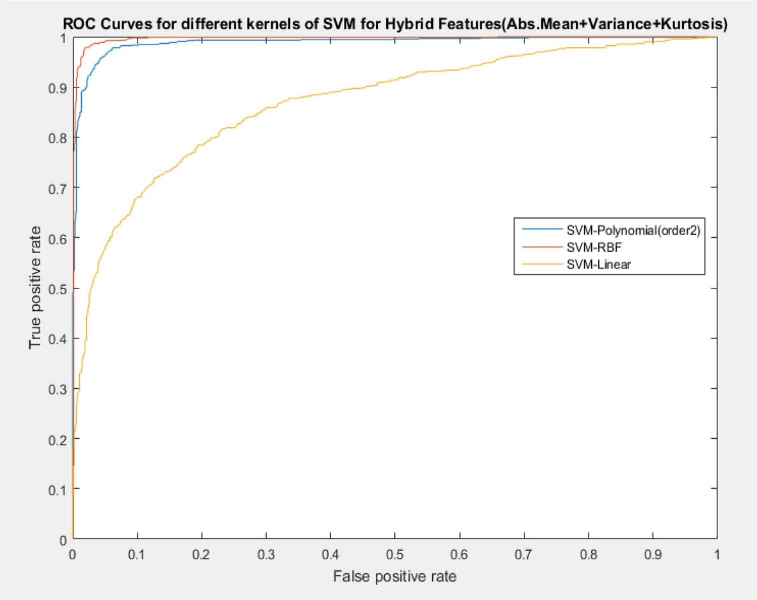


Fig. 11 ROC Curve for Hybrid feature vectors

Table 9 Classification Accuracy with Increasing number of features for Fisher Method

No. of features	Efficiency of SVM models (%)			No. of features	Efficiency of SVM models (%)		
	polynomial	Rbf	Linear		Polynomial	rbf	Linear
1	56.00	53.00	56.67	**26**	72.66	73.00	66.50
2	52.00	56.00	48.17	**27**	74.00	75.33	67.67
3	52.67	55.00	49.00	**28**	73.83	74.33	67.83
4	50.50	55.83	50.83	**29**	74.00	73.33	68.00
5	55.33	58.67	50.17	**30**	74.00	69.00	68.00
6	56.67	56.83	51.00	**31**	74.83	68.33	67.67
7	57.50	60.33	50.67	**32**	75.83	68.50	68.50
8	59.83	63.00	52.17	**33**	74.50	67.50	68.50
9	65.00	63.50	55.83	**34**	74.16	72.33	68.66
10	65.34	61.83	56.50	**35**	73.50	72.67	68.83
11	67.50	63.67	58.83	**36**	73.33	73.83	68.67
12	61.17	62.67	56.33	**37**	73.33	73.33	68.34
13	65.17	63.50	57.67	**38**	72.67	73.67	68.16
14	66.84	64.83	57.17	**39**	72.83	73.17	67.84
15	67.50	68.33	59.67	**40**	76.50	77.83	73.00
16	65.17	66.67	60.83	**41**	75.16	77.83	72.83
17	65.50	66.34	60.17	**42**	75.66	77.50	71.17
18	69.34	67.17	68.83	**43**	76.16	77.50	70.83
19	69.00	66.17	69.50	**44**	77.50	77.66	71.67
20	73.00	68.83	70.67	**45**	77.00	79.66	71.67
21	72.80	68.67	70.17	**46**	76.50	77.83	71.67
22	74.00	67.17	70.50	**47**	78.50	79.00	72.50
23	72.50	67.83	71.33	**48**	78.33	78.66	71.83
24	74.67	69.50	70.50	**49**	78.50	**80.34**	71.83
25	72.67	72.33	66.16	**50**	77.33	78.84	72.00

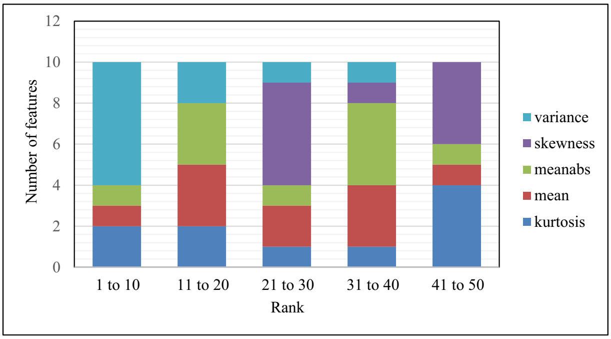


Fig. 12 Rank distribution of different features using Fisher Method

itself in much iteration as no new feature vector had F-ratio exceeding the threshold. Such rows were removed from Table 10 to avoid repetitions of data. Best results were seen for RBF kernel with k = 66 generating 38 feature vectors with an efficiency of 82.16%. As F-ratio for rest 11 columns had very diminishing value, they couldn't participate in performance evaluation.

Figure 14 plots the feature distribution for different values of *k*. A clear dominance of Variance and Absolute Mean can be seen. A total of 39 feature vectors are selected for k = 95. Out of these, 10 feature vectors from each of Variance and Absolute Mean, 8 feature vectors from Mean, 6 feature vectors from Kurtosis and 5 feature vectors from Skewness were selected respectively. Figure 15 plots the performance curve for SVM for best performing set of Hybrid Features.

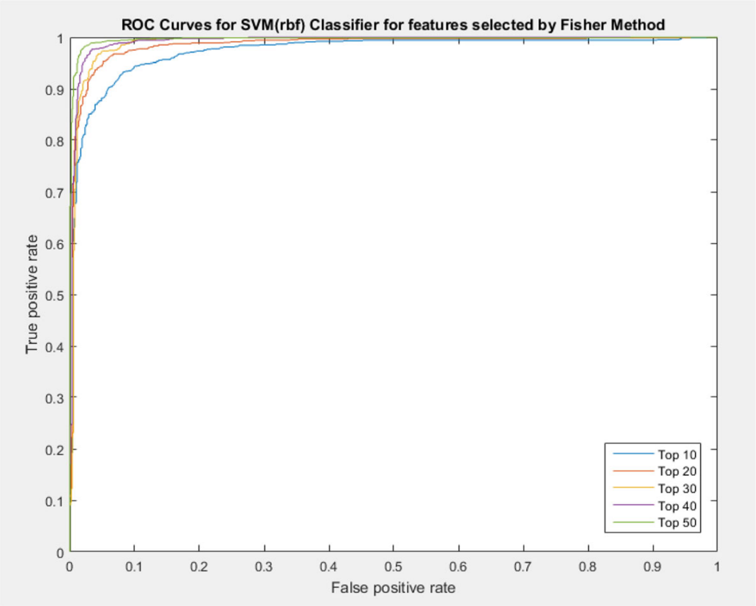


Fig. 13 ROC Curve for Hybrid feature vectors selected using Fisher Method

Table 10 Classification Accuracy for different threshold values of F-ratio

Sl. No.	k	Threshold	No. of Feature vectors selected	Efficiency of SVM models (%)		
				polynomial	Rbf	linear
1	1	0.033430347	1	52.17	52.83	50.00
2	2	0.016715173	7	57.50	60.00	50.67
3	3	0.011143449	9	62.17	61.83	53.67
4	4	0.008357587	12	61.33	68.83	57.50
5	5	0.006686069	13	66.50	69.50	59.17
6	6	0.005571724	15	73.17	70.83	64.67
7	7	0.004775764	16	70.83	69.50	64.50
8	8	0.004178793	17	70.67	69.67	64.50
9	9	0.003714483	19	72.17	69.66	70.17
10	11	0.003039122	21	70.33	70.00	67.83
11	14	0.002387882	22	71.00	69.83	67.17
12	15	0.00222869	23	72.33	71.00	68.17
13	18	0.001857241	26	72.50	74.16	69.33
14	19	0.001759492	28	75.00	73.50	68.50
15	20	0.001671517	29	75.16	73.33	68.50
16	22	0.001519561	30	74.67	69.67	69.00
17	23	0.001453493	31	75.67	69.33	69.00
18	24	0.001392931	32	74.33	72.67	69.00
19	26	0.001285783	33	74.50	66.83	68.50
20	32	0.001044698	33	75.33	66.83	68.50
21	40	0.000835759	34	74.17	73.83	67.83
22	43	0.00077745	35	71.83	73.83	68.16
23	58	0.000576385	36	78.16	78.00	70.66
24	61	0.000548038	37	78.67	80.50	73.67
25	66	0.00050652	38	78.33	**82.16**	73.00
26	95	0.000351898	39	79.83	81.50	73.33

7 Comparision with past work

The aim of the study was to investigate and understand the efficacy of EMD based statistical features in classifying a speech and low frequency music signal(guitar) sharing a common spectra range for different tuning parameters of state-of-the-art classifiers. This work doesn't propose a new set of features for efficient Speech Music Discrimination task rather focuses on analyzing the performance of EMD based features for classification task. Most of the works in

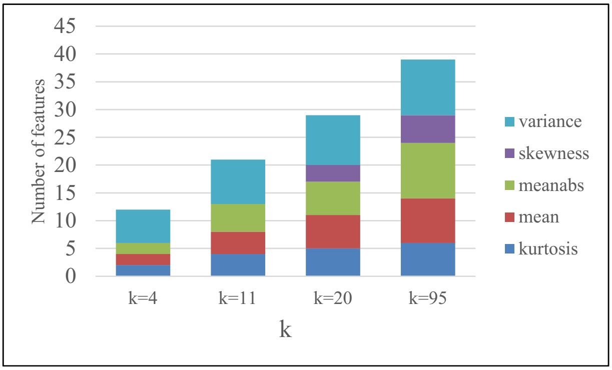


Fig. 14 Feature Distribution for different values of 'k'

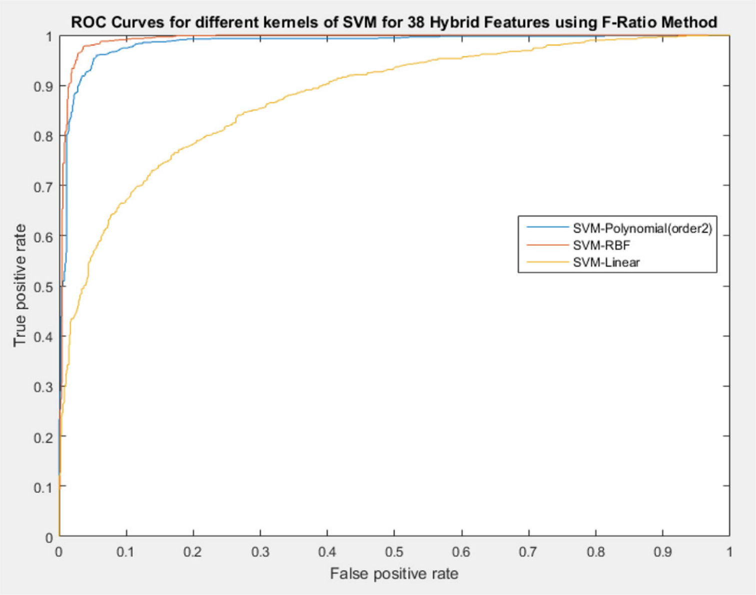


Fig. 15 ROC Curve for 38 Hybrid feature vectors selected using F-Ratio

literature are based on commercial speech and music segments either recorded from Television or Radio stations, with a wider spectrum spread. Hence, a comparison with other work would be futile. However, we compare and validate our results with a similar work done by Khonglah et al. where they explored statistical features from EMD to discriminate speech and music samples using S&S Database for broadcast Radio recordings [13]. They explored SVM and KNN in their work and achieved a highest accuracy of 90.83%. Following observations are tabulated in Table 11.

From Table 11, we conclude that classification of Speech and Low frequency music signal is more challenging than classification of speech and commercial music signals as even after

Table 11 Comparative Analysis

Sl. No.	Points of Comparison	Previous Work [13]	This Work
1	Speech Samples	From SS Database Training: 60 Samples Testing: 20 Samples Frequency Range: 0–4000 Hz	From SS Database Training: 60 Samples Testing: 20 Samples Frequency Range: 0–4000 Hz
2	Music Samples	From SS Database (No vocals) Training: 60 Samples Testing: 20 Samples Frequency Range: 0–10,000 Hz (approx.)	From [34] Training: 60 Samples Testing: 20 Samples Frequency Range: 80–1200 Hz
3	Best Accuracy	90.83% using SVM	82.16% using SVM
4	Rank of Features	Absolute Mean>Kurtosis > Variance > Skewness > Mean	Absolute Mean>Variance > Kurtosis > Mean>Skewness
5	Feature Selection	–	F-Ratio and F-score
6	Feature Importance	Line plot and scatter plot	Line plot, Scatter plot, ROC analysis, Histogram plot
7	Classifiers	SVM, KNN	SVM, KNN, ANN, NB with tuning parameters

applying feature selection to raw features, the classification accuracy is below the earlier reported work. However, the rank of features for both the works almost matches each other with Absolute Mean, Variance and Kurtosis being the top ranked features for both scenarios which also validates our experimental works. Results from this study may be extended to Indian instruments like Bansuri and Been which also share a similar spectrum spread.

8 Conclusion

This work analyzes the performance of statistical features extracted from EMD for the classification of speech and low frequency guitar signal. A signal was decomposed into 10 IMF's and each of the IMFs was framed into one second for feature extraction. The variations of the extracted features across different IMFs were studied using the line plot. The discriminatory evidences in them were further validated using scatter plot and ROC plot. Initial experiments were run on isolated features with four different classifiers. Absolute Mean and Variance stood as best performing features while SVM and ANN stood as the best classifiers. Best classification accuracy of 68.83% was observed for Absolute Mean feature when used with SVM (RBF). Further analysis was done on the discriminatory characteristics of hybrid features using scatter plot. Different hybrid features were created by combination of two or more isolated features. Overall improvement in performance was seen for all the four classifiers. Best results were observed for the hybrid of Absolute Mean, Variance and Kurtosis Features with an accuracy of 82.00%. An improvement of 19.13% is seen for best performing hybrid features over best performing isolated feature. To further validate the results, two different techniques for Feature Selection were evaluated on the dataset. Results from both Fisher Method and F-ratio indicated Variance and Absolute Mean as best performing features. Best efficiency of 82.16% is found with SVM classifier (rbf) with 38 features (10 Variance +10 Absolute Mean + 8 Mean + 5 Skewness +5 Kurtosis). Future work may concentrate on studying the application of EMD and its variant for analysis of polyphonic and folk music.

Acknowledgements We, all the authors, would like to thank Prof. Dr. Sandeep Singh Solanki, HOD and Professor, BIT Mesra, for his guidance and support in this research work. We would also like to extend my gratitude to Birla Insititute of Technology for providing us with the facilities to conduct our research.

References

1. Alexandre-Cortizo E, Rosa-Zurera M, Lopez-Ferreras F (2005) Application of fisher linear discriminant analysis to speech/music classification. EUROCON 2005 - The International Conference on "Computer as a Tool", pp 1666–1669. https://doi.org/10.1109/EURCON.2005.1630291
2. Babiker A, Faye I, Mumtaz W, Malik AS, Sato H (2018) EEG in classroom: EMD features to detect situational interest of students during learning. Multimedia Tools and Applications, pp:1–21
3. Birajdar GK, Patil MD, (2018) Speech and music classification using spectrogram based statistical descriptors and extreme learning machine. Multimedia tools and applications, pp.1-28.
4. Bouzid A, Ellouze N (2004) "Empirical mode decomposition of voiced speech signal," in Control, Communications and Signal Processing, 2004. First International Symposium on. IEEE, pp. 603–606
5. Bykhovsky D, Hadar O (2010) Evaluation of a GLRT threshold for voiced-unvoiced decision and pitch tracking in noisy speech. 2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel, pp 000680–000683. https://doi.org/10.1109/EEEI.2010.5662126

6. Flandrin P, Rilling G, Goncalves P (2004) Empirical mode decomposition as a filter bank. Signal Processing Letters, IEEE 11(2):112–114
7. Gu Q, Li Z, Han J, (2012) Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725. https://doi.org/10.48550/arXiv.1202.3725
8. Huang NE, (2014) Hilbert-Huang transform and its applications (Vol. 16). World scientific
9. Huang H, Pan J (2006) Speech pitch determination based on Hilbert Huang transform. Signal Process 86(4): 792–803
10. Huang NE, Shen SS (2005) Hilbert-Huang transform and its applications. World Scientific 5
11. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In proceedings of the Royal Society of London a: mathematical, physical and engineering sciences. R Soc 454(1971):903–995
12. Khonglah BK, Prasanna SM (2016) Speech/music classification using speech-specific features. Digital Signal Processing 48:71–83
13. Khonglah BK, Sharma R, Mahadeva Prasanna SR, (2015) Speech vs music discrimination using empirical mode decomposition. 2015 Twenty First National Conference on Communications (NCC), pp 1–6. https://doi.org/10.1109/NCC.2015.7084865
14. Kim SK, Chang JH (2009) Speech/music classification enhancement for 3GPP2 SMV codec based on support vector machine. IEICE Trans Fundam Electron Commun Comput Sci 92(2):630–632. https://doi.org/10.1587/transfun.E92.A.630
15. Lahmiri S, Gargour C, Gabrea M, (2012) Statistical features selection from intrinsic mode functions for pathologies detection in retina digital images. IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, pp 1585–1590. https://doi.org/10.1109/IECON.2012.6388532
16. Lim C, Chang JH (2015) Efficient implementation techniques of an svm-based speech/music classifier in SMV. Multimed Tools Appl 74(15):5375–5400
17. Moreno PJ, Rifkin R, (2000) Using the fisher kernel method for web audio classification. In acoustics, speech, and signal processing, 2000. ICASSP'00. Proceedings. 2000 IEEE international conference on (Vol. 4, pp. 2417-2420). IEEE
18. Panagiotakis C, Tziritas G (2002) A speech/music discriminator based on RMS and zero-crossings. 2002 11th European Signal Processing Conference, pp 1-4
19. Pantazis Y, Rosec O, Stylianou Y (2011) Adaptive AM–FM signal decomposition with application to speech analysis. IEEE Trans Audio Speech Lang Process 19(2):290–300
20. Papakostas M, Giannakopoulos T (2018) Speech-music discrimination using deep visual feature extractors. Expert Syst Appl 114:334–344
21. Roffo G, Melzi S, (2017) Ranking to learn: feature ranking and selection via eigenvector centrality. In new Frontiers in mining complex patterns: 5th international workshop, NFMCP 2016, held in conjunction with ECML-PKDD 2016, Riva del Garda, Italy, September 19, 2016, revised selected papers (Vol. 10312, p. 19). Springer
22. Ruiz-Reyes N, Vera-Candeas P, Muñoz JE, García-Galán S, Cañadas FJ (2009) New speech/music discrimination approach based on fundamental frequency estimation. Multimed Tools Appl 41(2):253–286
23. Sahoo JP, Ari S, Ghosh DK (2018) Hand gesture recognition using DWT and F-ratio based feature descriptor. IET Image Process 12(10):1780–1787
24. Saunders J, (1996) Real-time discrimination of broadcast speech/music. In ICASSP (pp. 993-996). IEEE
25. Scheirer E, Slaney M (1997) Construction and evaluation of a robust multi-feature speech/music discriminator. In Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on. IEEE 2:1331–1334
26. Seck M, Bimbot F, Zugaj D, Delyon B, (1999) Two-class signal segmentation for speech/music detection in audio tracks. In Sixth European Conference on Speech Communication and Technology
27. Sharma R, Prasanna SM, (2015) Characterizing glottal activity from speech using empirical mode decomposition. In communications (NCC), 2015 twenty first National Conference on (pp. 1-6). IEEE
28. Shirazi J, Ghaemmaghami S (2010) Improvement to speech-music discrimination using sinusoidal model based features. Multimed Tools Appl 50(2):415–435. https://doi.org/10.1007/s11042-009-0416-3
29. Tsipas N, Vrysis L, Dimoulas C, Papanikolaou G (2017) Efficient audio-driven multimedia indexing through similarity-based speech/music discrimination. Multimed Tools Appl 76(24):25603–25621. https://doi.org/10.1007/s11042-016-4315-0
30. Wang G, Chen XY, Qiao FL, Wu Z, Huang NE (2010) On intrinsic mode function. Adv Adapt Data Anal 2(03):277–293
31. Williams G, Ellis DP, (1999) Speech/music discrimination based on posterior probability features. Eurospeech 99: 6th European Conference on Speech Communication and Technology: Budapest, Hungary, September 5–9. https://doi.org/10.7916/D8KH0XRH

32. Wu Z, Huang NE (2004) A study of the characteristics of white noise using the empirical mode decomposition method. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences 460(2046):1597–1611
33. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(01):1–41. https://doi.org/10.1142/S1793536909000047
34. YouTube. (2019). Relaxing Music from Sungha Jung (The Best of). [Online] Available at: https://www.youtube.com/watch?v=IP8vBL5Q8Ac&t=338s. Accessed 05 Jan 2021
35. Zhang T, Kuo CCJ (2001) Audio content analysis for online audiovisual data segmentation and classification. IEEE Transactions on speech and audio processing 9(4):441–457