ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

Audio Type Identification Using EEMD: A Noise Assisted Data Analysis Method

Mr. Vinayak D. Chavan[1], Dr. Sanjay L. Nalbalwar[2]

M.tech student, Dept. of Electronics & Telecommunication Engg., Dr. B. A. T. U. Lonere, M.S., India[1]

Professor & Head Dept. of Electronics & Telecommunication Engg., Dr. B. A. T. U. Lonere, M.S., India[2]

ABSTRACT: Audio classification is a process of assigning particular class to an audio signal. Classifying the audio signal has many applications in the field of digital library, automatic organization of databases etc. In the last several years efforts have been made to develop different methods to extract information from audio signals, so that they may be stored, organized and retrieved automatically whenever required. In this work, audio signals are classified into different categories based on spectral and temporal features. In this methodology, the audio signal is initially decomposed into overlapped frames. Ensemble Empirical Mode Decomposition (EEMD), which is noise assisted data analysis method, is used to convert these frames into a set of band-limited functions known as Intrinsic Mode Functions (IMFs). Temporal and Spectral features then extracted from these IMFs and thereafter classification is done using Gaussian Mixture Model (GMM) classifier. Different combinations of features were tested to create feature vector. The experimental results showed accuracy of more than 80%.

KEYWORDS--Intrinsic mode function, ensemble empirical mode decomposition, spectral and temporal features, Gaussian Mixture Model.

I. INTRODUCTION

In the past decade a huge amount of multimedia data in the form of text, images, audio and video has become available. With the increasing use of such audio data and challenges faced in different multimedia application, it has become essential to put effort into audio signal analysis. An audio signal classification system should be able to categorize different audio input formats. Particularly, detecting the audio type of a signal (clean speech, speech with environmental background noise, and speech with music) allows such new applications as automatic organization of audio databases, segmentation of audio streams, intelligent signal analysis, intelligent audio coding, automatic bandwidth allocation, automatic equalization, automatic control of sound dynamics etc. All classification systems employ the extraction of a set of features from the input signal. In our method the features are extracted from the signal such that their discrimination capability is increased.

We organize the remainder of this paper as follows: Section 2 gives related work in this field. The proposed methodology is presented in section 3 where we have described the EMD and EEMD techniques. Section 4 describes the commonly used features for classification and the basic concepts of Gaussian mixture model classifier. The database used for the experiments and experimental results is presented in Section 5. We conclude in Section 6 with conclusions and future work in the field.

II. RELATED WORK

The five audio classes: silence, speech, music, speech with music and speech with noise is classified using feature extraction matrix in [1]. The classification of traffic noise sources: motorbikes, cars and heavy trucks are made in [2] by using spectral features like spectral centroid, spectral roll-off, sub-band energy ratio and zero-crossing rate as temporal feature. In [3] four types of background noise sources are classified using empirical mode decomposition method with accuracy up to 80-85%. Speech and music signals were classified using EMD in [7] with 85-90%

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

accuracy. Three feature sets for representing timbral texture, rhythmic content and pitch content of music signals were proposed and evaluated using statistical pattern recognition classifiers with 61% accuracy in [9] for classifying ten musical genres.

In this paper, we have proposed a method for audio type identification as clean speech, speech with noise and music which relies on temporal and spectral shape features. Noisy speech is again classified into five subcategories as airport, babble, station, train and street [3], [4]. Also music is again classified as rock, pop, jazz, and disco. In this experiment, the input signal is hierarchically decomposed by ensemble empirical mode decomposition (EEMD) [6], [7]. A given signal is decomposed into a number of Intrinsic Mode Functions (IMFs) and a residual. The features are computed from IMFs and the residue. The performance of various features extracted from the IMFs is evaluated using Gaussian mixture model classifier.

III. PROPOSED METHODOLOGY

2.1 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) is one of the methods of feature extraction. It is advantageous compared to the other methods as EMD is an adaptive data analysis method that is based on local characteristics of the data, and hence, it catches nonlinear, non-stationary oscillations more effectively. EMD method is able to decompose a complex signal into a series of intrinsic mode functions (IMF) and a residue. [6], [7].

2.1.1 Intrinsic Mode Function

The EMD decomposes the original signal into a definable set of adaptive basis of functions called the intrinsic mode functions. Each IMF must satisfy two basic conditions: In the whole data set, the number of local minima or local maxima and the number of zero crossings must be equal or they may differ at most by one. At any point, the mean value of the envelope, one defined by the upper envelope and the other by the lower envelope is zero. To generate the IMFs sifting process is applied. The signal is decomposed until we get the final component as residue.

1. identify all extrema of x(t).
2. Interpolate the local maxima to form an upper envelope u(t).
3. Interpolate the local minima to form an lower envelope l(t).
4. Calculate the mean envelope: $m(t)=[u(t)+l(t)]/2$.
5. Extract the mean from the signal: $h(t)=x(t)-m(t)$
6. if h(t) satisfies the IMF Condition, stop sifting else keep sifting.

2.2 Ensemble Empirical Mode Decomposition

EMD is a dyadic filter bank for any white noise-only series, when the data is intermittent; the dyadic property is often compromised. Hence, adding noise to the data could provide a uniformly distributed reference scale. This enables EMD to repair the compromised dyadic property; and the corresponding IMFs of different series of noise have no correlation with each other. Therefore, the means of the corresponding IMFs of different white noise series are likely to cancel each other [8]. With these properties of the EMD in mind, the EEMD is developed as follows:

1. Add a white noise series to the targeted data.
2. Decompose the data with added white noise into IMFs.
3. Repeat step 1 and step 2 again and again, but with different white noise series each time.
4. Obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result.

IV. FEATURE EXTRACTION AND CLASSIFICATION

Features extraction is the main step for classifying any audio signal into a given class [5]. These features will decide the class of the signal. Feature extraction involves the analysis of the input signal. The feature extraction techniques can be classified as temporal analysis and spectral analysis technique. Temporal analysis uses the waveform of the audio signal itself for analysis. Temporal features include Short Time Autocorrelation Function (ACF), Short Time Energy

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

(STE), and Zero Crossing Rate (ZCR). Spectral analysis utilizes spectral representation of the music signal for analysis. Spectral features include Spectral Centroid (SC), Spectral Roll off (SR) and Spectral Flux (SF) [9]. Using these features, feature vector is formed which is then applied to the classifier for classification. In our experiment we used Gaussian Mixture Model classifier with Maximum Likelihood.

3.1 Gaussian Mixture Models (GMM)

GMMs model the distribution of feature vectors. For each class, we assume the existence of a probability density function expressible as a mixture of a number of multidimensional Gaussian distributions. The iterative Expectation Maximization (EM) algorithm is usually used to estimate the parameters for each Gaussian component and the mixture weights. GMMs have been widely used as classifiers by using a maximum likelihood criterion to find the model best suited to a particular audio [6].

V. EXPERIMENTS AND RESULTS

In out experiment, we have classified audio signal into three major classes: clean speech, speech with background noise and music. Then the noisy speech signal is again classified into five sub-categories: airport, babble, station, train and street. Music signal is classified into four sub-categories: pop, rock, jazz and disco. For this experiment we selected dataset consisting of 30 clean speech signals, 150 noisy speech signals and 120 music signals of 2-5 sec duration long. The noisy speech consists of one type of noise mixed with speech with different utterances. The given signal is first segmented into frames of 50 ms duration and 25 ms overlap between each frame. Then each frame is decomposed into IMFs using Ensemble Empirical Mode Decomposition. These IMFs are then used for extracting different features as mentioned in section 3. If there is 'N' no. of frames for one signal then mean of particular feature of all the frames is considered as one of the feature for that signal.

$$Fr = (1/N)\sum_{i=1}^{N} fr(i) \qquad (1)\ ;$$

where 'r' denotes feature no. and 'i' denotes the frame no. of one particular signal. From these features we formed the feature vector which is then applied for classification.

$$Fv, m = [F_1\ F_2\ F_3 \ldots .\ F_r] \qquad (2)$$

Where 'v' is signal no., 'm' in IMF no. and 'r' is feature no.

Feature vectors are formed by combining different set of features for each audio category and then Classification is performed using GMM classifiers with Maximum Likelihood. Results of classification which are having very good accuracy are presented below. Others are not presented as the accuracy was not so good.

Table -1: Classification accuracy for major audio classes IMF1 (%)

FEATURES	SPEECH	SPECH + NOISE	MUSIC	OVERALL ACCURACY
ZCR	100	100	93.33	97.78
STE	86.67	46.67	100	77.78
ZCR,MFCC	100	100	100	100
SR,SC,SF	100	93.33	100	97.78
ZCR,STE	100	100	100	100
ZCR,SR,SC,SF	100	100	100	100

From the above results it is clear that, individual features like ZCR, STE can discriminate between speech and music. All music samples are correctly identified. But between speech and speech with noise identification, the accuracy is quite less. So, different feature combinations were tested and observed that the accuracy up to 100% can be achieved.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

Table - 2: Classification accuracy for music genre IMF1 (%)

FEATURES	POP	ROCK	JAZZ	DISCO	OVERALL ACCURACY
SR,SC,SF,STE	100	93.33	66.67	40	75
ZCR,SR,STE,MFCC	100	93.33	93.33	60	86.67
ZCR, SR,MFCC	100	93.33	86.67	73.33	88.67
ZCR,SF,MFCC	100	86.67	100	66.67	88.33
ZCR,SR,MFCC	100	93.33	86.67	73.33	88.67
ZCR,SR,SC,MFCC	100	100	86.67	73.33	90
ZCR,SR,SF,MFCC	100	93.33	93.33	80	91.67
ZCR,SR,SC,SF,MFCC	100	100	93.33	73.33	91.67

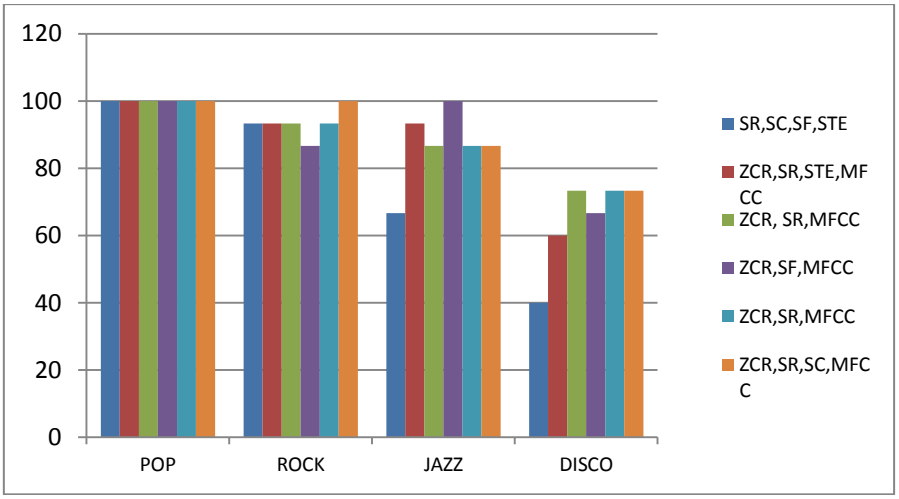

Fig. 1: bar chart showing accuracy of the feature combinations

Music signals with different genre are highly similar. Hence, individual features are not sufficient to classify all the genres. From above results, it is clearly observed that the obtained accuracy is up to 92% which is very good.

Table - 3: Classification accuracy for background noise IMF1 (%)

FEATURES	AIRPORT	BABBLE	STATION	TRAIN	STREET	OVERALL ACCURACY
ZCR,SR,SC,SF	86.67	100	100	93.33	100	96
ZCR,SR,SC,SF,STE	86.67	100	100	93.33	100	96
ZCR, SR, SC, SF, STE, MFCC	100	100	93.33	93.33	86.67	94.67
SR,SC,SF	86.67	100	93.33	93.33	86.67	92
ZCR,SC,SF,STE	100	80	100	100	80	92

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

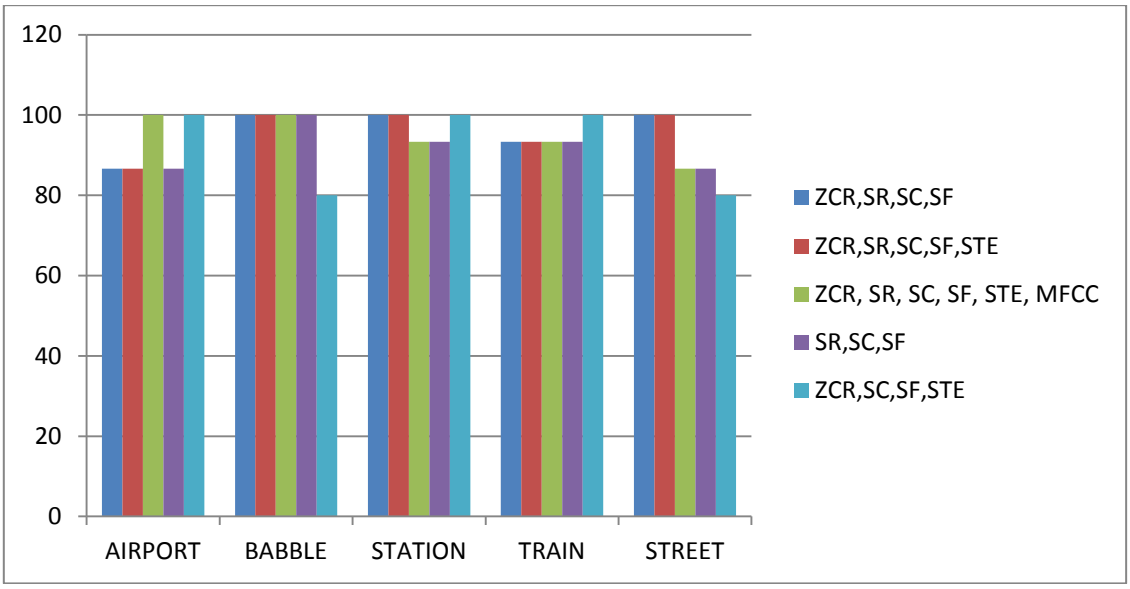

Fig 2: bar chart for the accuracy of the feature combinations for music genre

Noise signal classification is important in case speaker/speech recognition. As observed in table 3, background noise sources, with feature combinations, can be classified with more than 90% accuracy, which is very good.
Also it is observed that first IMF can discriminate between the specified categories with good accuracy. Other IMFs also tested for classification but the accuracy is less than the accuracy of IMF1. Hence results for other IMFs are not mentioned.

VI. CONCLUSION

Automatic audio classification is a complicated and problematic task, but still has important value both in research and commercialized applications. In this paper, we introduced use of EEMD for different types audio classification by decomposing the signals into no. of IMFs and extracting spectral and temporal features from these IMFs. The values of selected features are used to form feature vector by combining different features and then classification is done. Experimental results show that EEMD can be successfully used for different types of audio classification whether it is speech/music classification, environmental noise classification or music genre classification with accuracy more than 85%. For future work, different kinds of features can be extracted and various classifiers can also be used to improve the accuracy. Also audio signals of longer duration can be used to get more accurate classification.

REFERENCES

[1] N. Nitanda, M. Haseyama, and H. Kitajima, "Accurate audio segment classification using feature extraction matrix," *in Proc. ICASSP, 2005*
[2] M. A. S. Seoane, A. R. Molares, and J. L. A. Castro, "Automatic classification of traffic noise," *in Proc. Acoustics'08, Paris, June 29–July 4, 2008*
[3] Deepak Jhanwar, Kamlesh K. Sharma and S. G. Modani, "Classification of Environmental Background Noise Sources Using Hilbert-Huang Transform", International Journal of Signal Processing Systems Vol. 1, No. 1 June 2013
[4] B.Han and E. Hwang, "Environmental sound classification based on feature collaboration," *in Proc. ICME, 2009*
[5] T. Lidy, R. Mayer, A. Rauber, P. J. Ponce de Leon, A. Pe rtusa, and J. M. Inesta, "A Cartesian Ensemble of Feature Subspace Classifiers for Music Categorization", *11th International Society for Music Information Retrieval Conference (ISMIR 2010), 2010, pp. 279-284.*
[6] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen,C. Tung, and H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Of the Royal Society of London , vol. 454, pp. 903–995, 1998.*
[7] Arijit Ghosal, Bibhas Chandra Dhara, Sanjoy Kumar Saha, "Speech/Music Classification Using Empirical Mode Decomposition", *Second International Conference on Emerging Applications of Information Technology, 2011.*
[8] Zhaohua Wu, Norden E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method", *Advances In Adaptive Data Analysisvol. 1, No. 1 (2009) 1–41*
[9] George Tzanetakis, Perry Cook, "Musical Genre Classification of Audio Signals", *Ieee Transactions On Speech And Audio Processing, Vol. 10, No. 5, July 2002*

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2014

[10] Sujeet Kini, Sankalp Gulati, and Preeti Rao, "Automatic Genre Classification of North Indian Devotional Music", *National Conference on Communications, 2011, pp. 16-20*.
[11] M.J. Carey, E. S. Parris, and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination," *in ICASSP, April 1999*
[12] Megha Agarwal, R.C.Jain, "Ensemble Empirical Mode Decomposition: An adaptive method for noise reduction ", *IOSR-JECE e-ISSN: 2278-2834,p- ISSN: 2278-8735. Volume 5, Issue 5 (Mar. - Apr. 2013), PP 60-65*
[13] http://homepages.cae.wisc.edu/~ece539/matlab/
[14]]https://github.com/marsyas/marsyas/tree/master/collections
[15] http://ecs.utdallas.edu/loizou/speech/noizeus/