Classification of Environmental Background Noise Sources Using Hilbert-Huang Transform

Deepak Jhanwar
Govt Engineering College Ajmer/ECE Department, Ajmer, India
deepakjhanwar2001@gmail.com

Kamlesh K. Sharma and S. G. Modani
Malaviya National Institute of Technology, Jaipur /ECE Department, Jaipur, India
kksharma_mrec@yahoo.com, shrigmodani@gmail.com

***Abstract*—Classification of common environmental background noise sources like train, airport, car and exhibition hall mixed with speech signals comprise of different utterances of various speakers required in many applications especially for forensic purposes. The acoustic features based on the Fourier and wavelet transformations are frequently used for environmental noise classification purpose. In the methodology presented, the noisy speech signal under test is initially decomposed into overlapped frames. In order to convert the latter into a set of band-limited functions known as Intrinsic Mode Functions (IMFs), we use Empirical Mode Decomposition (EMD) method through Hilbert-Huang Transform (HHT). In this method the spectral and temporal features are extracted from the IMFs of different noisy speech signals and thereafter classification of different noise sources has been done using k-Nearest Neighbor (k-NN) classifier as a simpler way. Computed individual feature provides success rate of discrimination which varies from 77% to 85%. The combined feature vector enhances the success rate of classification as exhibited in the results. The technique presented reduces the dimensions of feature vector and proves the effectiveness of individual feature extracted from IMFs for the same.**

***Index Terms*—Hilbert-huang transform, intrinsic mode function, empirical mode decomposition, k-nearest neighbor classifier**

I. Introduction

The sound contains a unique signature of corresponding source. If right characteristics are taken into consideration, temporal and spectral characteristics provide suitable information regarding the type of source. The presented work proposes new feature set to discriminate different types of environmental background noise sources irrespective of speakers and utterances to understand the surrounding environment of the speaker.

It is an important research field and lots of efforts are being made in this direction. Ref. [1], the ten sound categories are classified on the basis of examining Motion Picture Expert Group (MPEG-7) audio standards. The low level descriptor audio spectrum projection is discussed and added to Hidden Markov Model (HMM). Conventional Maximum Log Likelihood (MLL) estimation is used to assign class to all the samples under test. Ref. [2], Mel-Frequency Cepstral Coefficients (MFCCs) are added in parallel with basis decomposition methods and demonstrates better performance with practical constraints. The ten audio classes are recognized by temporal integration of features in [3]. Ref. [4], the features used to detect human voice are zero-crossing rate, low short time energy ratio, spectrum flux and poles and zeroes of Linear Prediction Coefficients (LPC) model. The technique of classifying heart sound is presented in [5] using S-transform in addition with multilayer perceptron neural network classifier. The method of discriminating different environmental background noise sources like train, airport car, restaurant and rain is elaborated in [6] and autocorrelation functions based physical parameters are used for the same. The non-speech natural sounds are classified by using product code HMM models. In this scheme as explained in [7], two HMM's used per class, one for spectral shape and other for gain. Decisions on classification are made by scoring sequences of shape and gain indexes from a product code Vector Quantization (VQ). Ref. [8], a rough set based feature selection method is presented. The key idea of this technique is that most effective features can discriminate the large number of samples belonging to various classes of noise sources. A Noise Sources Automatic Classification System (NSACS) has been developed. The five audio classes: silence, speech, music, speech with music and speech with noise background is classified using feature extraction matrix as discussed in [9]. The classification of traffic noise sources: motorbikes, cars and heavy trucks are made in [10] by using spectral features like spectral centroid, spectral roll-off, sub-band energy ratio and zero-crossing rate as temporal feature in addition with basic classifiers. The feature collaboration method of environmental sound classification is discussed in [11]. The fuzzy pattern classification with new results and pitch based classification method are discussed in [12] [13].

The acoustic features based on Fourier and wavelet transforms are common and are used frequently for environmental sound classification purpose but the common features based on HHT in analyzing background

Manuscript received May 4, 2013; revised August 10, 2013

doi: 10.12720/ijsps.1.1.68-73

sound sources are proved to be much discriminating. The HHT add another time series that has been phase shifted by 90° and contributes a sifting process that decomposes the signal into a set of band-limited functions known as IMF. The spectral and temporal features like cross-correlations, spectral centroid, spectral flux, short time energy, zero-crossing rate, maximum incoherence index, MFCC of decomposed IMFs corresponding to different signals has been evaluated after pre-processing of signals These features are found helpful and efficient in discriminating the different environmental background sound sources independent of speakers and utterances. Common environmental sound sources like train, airport, car and exhibition are considered here as background noise. Finally simple k-NN classifier is used to classify different background sources.

The work presented is distributed in different sections as follows: Section II shows the overview of HHT. The description of proposed methodology is discussed in section III. The experimental work on MATLAB® platform is presented in section IV. Finally section V contains the conclusion of the paper.

II. Overview of HHT

The HHT developed by Huang in his work [14]-[16] proved to be useful in analyzing the non-stationary and nonlinear acoustic signals. The HHT contains EMD as one part and Hilbert Spectral Analysis (HSA) as another part. The physical characteristics of acoustic signals are carried by instantaneous frequency which can be computed easily by Hilbert transform.

The EMD method is required to deal with non-stationary and nonlinear processes. The decomposition is based on the assumption that any non-stationary and nonlinear signal consists of different intrinsic modes of oscillations. Each intrinsic modes, whether linear or nonlinear, shows a simple oscillation, consisting of the same number of extrema and zero-crossings. The oscillation will also be symmetric with respect to the local mean. At any given time, the signal to be analyzed may have many different coexisting and superimposed modes of oscillation and each of these oscillatory modes is represented by an IMF as follows:

- In the whole signal, the number of maxima or
- minima and the number of zero-crossings must either be equal or differ at the most by one, and
- At any point, the mean value of the envelope defined either by local maxima or by the local minima is zero.

An IMF represents a simple oscillatory mode which can have a variable amplitude and frequency as functions of time instead of constant amplitude and frequency, as in a simple harmonic component. With the definition for the IMF as mentioned, decomposition of any function can take place which is as follows:

- To identify all the local extrema of the given signal.
- Then connect all the local maxima by a cubic spline line to produce upper envelope.
- Repeat second step to produce the lower envelope by joining local minima.
- The upper and lower envelopes should cover the entire signal between them.

Let the signal be represented by $x(t)$,
Mean of upper and lower envelope is denoted by m.
First component,

$$p_1 = x(t) - m_1 \tag{1}$$

In the subsequent sifting process, p_1 can be treated as proto-IMF whereas in next step it is assumed as signal.

Next step,

$$p_2 = p_1 - m_2 \tag{2}$$

After repeated siftings

$$p_k = p_{k-1} - m_k \tag{3}$$

where p_k is the first $IMF(C_1)$.

Here, a decision to stop the process is critical. This criterion of stoppage is decided by Cauchy type of convergence test defined as,

$$SQD_k = \frac{\sum_{t=0}^{T} | p_{k-1} - p_k |^2}{\sum_{t=0}^{T} {p_{k-1}}^2} \tag{4}$$

If SQD_k is less than predetermined value then sifting process will be stopped. But this criterion does not satisfy the characteristics of IMF hence an S-number is selected up to which the number of zero-crossings and maxima remains equal or differ by one. After selecting S-number, the first $IMF = C_1$ is found,

$$res_1 = x(t) - C_1 \tag{5}$$

res_1 is residue which is treated as new signal containing longer period variations. The sifting process will be finished when residue becomes small and monotonic function from which *IMF* cannot be extracted. Let, the final residue corresponding to signal is res_N then,

$$x(t) = \sum_{n=1}^{N} C_n + res_N \tag{6}$$

where C_n shows 'n' IMFs.

III. Proposed Methodology

The methodology steps of proposed discrimination technique are as follows,

A. IMF Decomposition of Frames

The different noisy speech signals of common background noise are indicated by s=1, 2, 3,…

- The acquainted noisy speech signals are decomposed into frames with frame period of 50ms and overlap period of 25ms.

$$x(t) = \sum_{Frame,\, fr=1}^{F} X_{fr} \tag{7}$$

- Silent periods at the start and end of the frame are removed.
- Each frame is decomposed into corresponding IMFs till the residual is obtained.

$$X_{fr} = \sum_{n=1}^{N^{s}_{fr}} C_n + res_{N^{s}_{fr}} \tag{8}$$

where N^{s}_{fr} is nos. of IMFs corresponding to particular frame $(f = 1, 2, 3, ...)$ of noisy speech signal $(s = 1, 2, 3, ...)$

- The minimum value of N^{s}_{fr} is obtained out of all frames of the sample.

$$\text{Min. value of } N^{s}_{fr} = N^{s}_{\min} \tag{9}$$

where, $N^{s}_{\min}$ is an integer number.

B. Feature Vector Computation

- The particular feature value is computed from an individual IMF of the frame of corresponding noisy speech signal.
- Mean of the feature values of particular order of IMF of all the frames are computed and treated as feature value of corresponding signal for further computation.
- The feature vector contains the feature values corresponding to different IMFs of corresponding noisy speech signal.

$$feature_{vector_f} = \left[feature_{f(IMF-1)}, feature_{f(IMF-2)}, ... \right] \tag{10}$$

where $f = 1, 2, 3, ...$ indicate the type of feature, $feature_{f(IMF-N)}$ shows the feature value/ element of the corresponding $feature_{vector_f}$ and $N = 1, 2, 3, ... N^{s}_{\min}$ indicate the number of IMF.

- k-NN (k=3) classifier is used as simple classifier for the $feature_{vector_f}$ containing more than one element. The classifier is trained by 15 samples of each type of background noise and 20 samples per noise type are tested to produce results of classification.
- Classification of single element based $feature_{vector_f}$ is done on the basis of Euclidean distance computation between the feature value of the sample under test and the mean of the corresponding feature values of training samples for that particular noise.
- $combined\_feature_{vector_f}$ is also formed by collecting those values of different $feature_{vector_f}$ having maximum discrimination between different noise sources.

IV. Simulation Results and Analysis

A. NOIZEUS Noisy Speech Database

NOIZEUS is a noisy speech corpus recorded in laboratory to facilitate comparison of speech enhancement algorithms among research groups. The noisy database, corrupted by eight different real-world noises at different SNR, contains 30 IEEE sentences produced by three different male and three female speakers. The noise was taken from the AURORA database containing train (suburban), car, exhibition hall and airport noise. Thirty sentences are taken from the IEEE sentence database as these sentences are phonetically balanced with relatively low word-context predictability and recorded in a sound proof chamber using Tucker Davis Technologies (TDT) recording equipments. The sentences were originally sampled at 25 KHz and down sampled to 8 kHz. The noise signals were added to the clean speech signal at SNR of 5 dB, 10dB and 15dB.

B. Feature Values of Selected IMFs

Fig. 1 shows the different IMF plots of a speech signal with utterance “Men strive but seldom get rich”. mixed with background train noise with 5dB SNR.

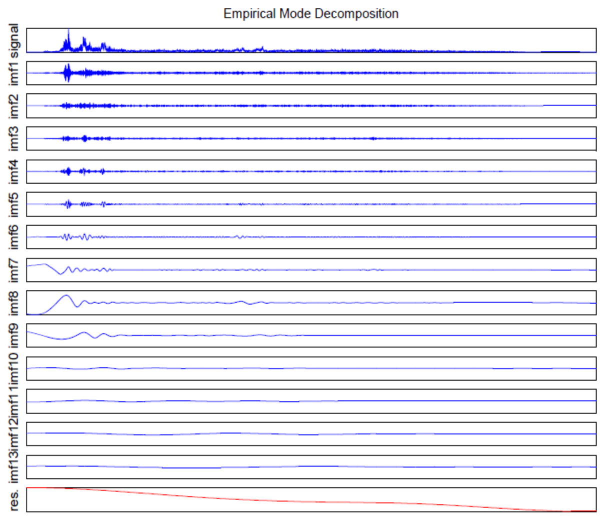

Figure 1. IMF plots of background noisy speech signal (train noise)

The different feature values indicated in Table I-VIII shows mean of the set of values of 10 noisy speech samples of each type of background noise source. Only those IMFs are selected and displayed in the corresponding tables which posses feature values of:

- Low variance across the values of various samples with same background noise
- Considerable discrimination among different background noise with same speech samples.
- All the individual features are classified by using k-NN classifier with k=3 and corresponding results are shown in the respective tables.

TABLE I. AUTOCORRELATION OF IMFS

Background Noise Sources	Mean Autocorrelation ($X10^{-3}$) (IMF-1)	Mean Autocorrelation ($X10^{-3}$) (IMF-2)	Success rate of classification by k-NN classifier
Train	29.1	53.3	85%
Airport	32.4	66.0	
Car	27.1	61.4	
Exhibition	39.5	57.3	

TABLE II. CROSS-CORRELATION BETWEEN IMFS

Background Noise Sources	Mean Cross-Correlation ($X10^{-3}$) (between IMF-1 & IMF-2)	Success rate of classification by computing Euclidean distance from mean value
Train	7.0	82%
Airport	11.2	
Car	9.0	
Exhibition	7.9	

TABLE III. INSTANTANEOUS FREQUENCY

Background Noise sources	IMF-1	IMF-2	IMF-3	Success rate of classification by k-NN classifier
Train	2.47	1.24	0.64	80%
Airport	2.00	0.99	0.53	
Car	2.20	1.08	0.58	
Exhibition	2.30	1.17	0.63	

TABLE IV. AVERAGE ZERO-CROSSING RATE

Background Noise sources	IMF-1	IMF-2	IMF-3	Success rate of classification by k-NN classifier
Train	1.04	0.43	0.21	77%
Airport	0.82	0.32	0.15	
Car	0.92	0.37	0.17	
Exhibition	0.96	0.40	0.19	

TABLE V. SPECTRAL CENTROID

Background Noise Sources	IMF-1 ($X10^{-2}$)	IMF-2 ($X10^{-2}$)	IMF-3 $X10^{-2}$	Success rate of Classification by k-NN classifier
Train	50.8	22.9	13.0	83%
Airport	40.7	18.7	11.4	
Car	44.2	20.3	12.3	
Exhibition	46.5	21.8	12.9	

TABLE VI. SPECTRAL FLUX

Background Noise Sources	IMF-1	IMF-2	IMF-3	Success rate of classification by k-NN classifier
Train	3.50	3.01	1.47	85%
Airport	3.60	2.01	1.28	
Car	4.70	2.18	1.45	
Exhibition	3.64	2.59	1.38	

TABLE VII. MFCC (MAX.)

Background noise sources	IMF-1	Success rate of classification by computing Euclidean distance from mean value
Train	4.62	78%
Airport	3.58	
Car	3.67	
Exhibition	4.11	

TABLE VIII. DELTA (MAX.)

Background Noise sources	IMF-1 ($X10^{-2}$)	IMF-2 ($X10^{-2}$)	IMF-3 ($X10^{-2}$)	Success rate of classification by k-NN classifier
Train	11.0	25.6	21.7	81%
Airport	13.16	34.5	17.6	
Car	11.7	34.0	20.0	
Exhibition	12.0	28.5	19.6	

The feature values corresponding to other IMFs are discarded due to higher variance for the samples of same type of background noise and negligible discrimination among different noise sources.

Table IX shows the change in maximum bicoherence index from one IMF to next IMF. Only those pairs of IMFs are indicated in which the change is either increasing or decreasing corresponding to particular background noise.

TABLE IX. MAXIMUM BICOHERENCE INDEX

Background Noise Sources	Common change in value of MBI observed between consecutive order of IMFs	Type of change (By more than 30%)	Success rate of classification (by matching results)
Train	1 to 2 and 3 to 4	Increase	90%
Airport	1 to 2 2 to 3	Decrease Increase	
Car	2 to 3 and 3 to 4	Increase	
Exhibition	2 to 3	Increase	

TABLE X. EFFECTIVE DISCRIMINATING FEATURES BETWEEN DIFFERENT PAIRS OF NOISE SOURCES

Background Noise sources	Airport	Car	Exhibition
Train	mean Autocorrelation (IMF-2), mean Cross-Correlation (between IMF-1 & IMF-2), Instantaneous frequency (IMF-1), Spectral centroid (IMF-1), Change in Max. Bicoherence Index, MFCC(IMF-1), Deltas(IMF-3)	mean Autocorrelation (IMF-2), Spectral flux (IMF-1)	mean Autocorrelation (IMF-1),
Airport	*	Spectral flux(IMF-1), Deltas(IMF-3)	mean Autocorrelation (IMF-1), mean Autocorrelation (IMF-2), MFCC (IMF-1) Deltas(IMF-2)
Car	*	*	mean Autocorrelation (IMF-1) Deltas(IMF-2) Spectral flux(IMF-1)
Combined feature vector	89% (excluding Max. Bicoherence Index)		

V. CONCLUSIONS

Discrimination of acoustic noise sources is cumbersome. The presented work is an effort to display the utility of Hilbert-Huang transformation in discriminating the different acoustic signals in the form of environmental background noise by decomposing the signals in terms of IMFs and extracting the temporal and spectral features from different IMFs. The values of selected features of different IMFs for different noise sources are analyzed and found that particular feature corresponding to specific IMF discriminates the noise sources from each other in corresponding set as compared to other IMFs for the same feature. In Table I, Table III-VI and Table VIII, the feature values for different IMFs forms feature vectors. These vectors provide success rate of classification from 77% to 85% with the help of K-NN Classifier. Table II and Table VII depict single element feature vectors which provide success rate of classification with the help of Euclidean distance computation which is 82% and 78% respectively. Table IX contains the trends of change in the value of maximum bicoherence index between consecutive IMFs and shows clear discrimination between train and airport noise. The presented technique reduces the number of features and increases the discrimination potential of individual features. Table X displays the most effective feature values from discrimination point of view and forms a combined feature vector. The success rate of classification is improved with the help of combined feature vector and k-NN classifier. For future work, lot of other suitable features may be extracted from corresponding IMFs and different classification techniques may also be applied for producing better results of discrimination acoustic noise sources for medical, industrial and other related applications.

REFERENCES

[1] M. Casey, "MPEG-7 sound recognition tools," *IEEE Trans. Cir. and Sys. for Video Tech.*, vol. 11, no. 6, pp. 737-747, June 2001.

[2] H. G. Kim and T. Sikora, "Comparison of audio spectrum projection features and mfcc applied to speaker recognition, sound classification and audio segmentation*,"* in *Proc. ICASSP*, Montreal, 2004, pp. 925-928.

[3] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "Sound classification based on temporal feature integration," in *Proc. 4th International Symposium on Communications Control and Signal Processing,* Cyprus, 2010.

[4] L. Lu, H. Jiang, and H.J. Zhang, "A robust audio classification and segmentation method," in *Proc. 9th ACM International Conference on Multimedia,* Ottawa, Canada, pp. 203-211.

[5] H. M. Hadi, M. Y. Mashor, M. Z. Suboh, and M. Sapawi, "Classification of heart sound based on s-transform and neural network," in *Proc. 10th International Conference on Information Science, Signal Processing and their Application,* 2010.

[6] J. Yang and B.D. Burkana, "The acoustic properties of different noise sources," in *Proc. 6th International Conference on Information Technology: New Generations*, 2009.

[7] J. P. Woodard, "Modeling and classification of natural sounds by product code hidden markov models," *IEEE Trans. Sig. Proc.,* vol. 40, no. 3, July 1992.

[8] X. Zeng and Y. Zhan, "Development of a noise sources classification system based on new method for feature selection," *Applied Acoustics*, vol. 66, pp. 1196-1205, 2005.

[9] N. Nitanda, M. Haseyama, and H. Kitajima, "Accurate audio-segment classification using feature extraction matrix," in *Proc. ICASSP*, 2005.

[10] 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.

[11] B. Han and E. Hwang, "Environmental sound classification based on feature collaboration," in *Proc. ICME,* 2009.

[12] F. Beritelli, S. Casale, and G. Ruggeri, "New result in fuzzy pattern classification of background noise," in *Proc. ICSP,* 2000.

[13] A. B. Nielsen, L. K. Hansen, and U. Kjems, "Pitch based sound classification," in *Proc. ICASSP,* 2006.

[14] N. E. Huang, Z. Shen, and S. R. Long *et al*. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." in *Proc. R.Soc. Lond. A*, 1998, pp. 903–995.

[15] N. E. Huang, "Introduction to Hilbert-Huang Transform and Its Related Mathematical Problems," in *Hilbert-Huang Transform*

and Its Applications, Huang N. E., Shen S. S. P., Ed, World Scientific, 2005, pp. 1-26.

[16] N. E. Huang, M. L. Wu, S. R. Long, S. S. Shen, W. D. Qu, P. Gloersen, and K. L. Fan, "A confidence limit for the empirical mode decomposition and hilbert spectral analysis," in *Proc. Royal Soc. Of Lond. A*, vol. 459, pp. 2317-2345.

Deepak Jhanwar received the B.E. (honors) degree in Electronics and Communication Engineering from M.B.M. Engineering College, Jodhpur, India, in 1994. Currently, he is pursuing Ph.D. with Malaviya National Institute of Technology, Jaipur, India in the field of background sound discrimination. Presently he is working as Assistant Professor with Govt. Engineering College, Ajmer, India in Department of Electronics and Communication Engineering. He joined the Bhabha Atomic Research Centre, Mumbai, India in 1994 as Scientific Officer (C). from 1997 to 2000, he worked in reputed industries in India as an electronic engineer. From 2000 onwards till now, he is engaged with teaching job.

Mr. Jhanwar is a life member of I.E.T.E., India.

Kamlesh Kumar Sharma received the B.E. (honors), and M.E. (honors) degrees in Electronics and Communication Engineering from Malaviya National Institute of Technology, Jaipur, India, in 1990 and 2001, respectively. He received the Ph.D. degree from Indian Institute of Technology, New Delhi, India. Presently, he is working as Associate Professor with the Malaviya National Institute of Technology. From 1990 to 1997, he served as a Lecturer with the Electronics and Communication Engineering Department, Malaviya National Institute of Technology. He joined the Indian Telecommunication Service for a year in 1997 as an Assistant Divisional Engineer Telecom. From 1998 to 2003, he again served as a Lecturer with Malaviya National Institute of Technology. He was deputed to the Indian Institute of Technology, New Delhi, India, from 2003 to 2006, for pursuing the Ph.D. degree under the quality improvement program of the Government of India. His research interests include sampling theory of signals, image processing, and fractional transforms. Currently he is supervising more than 15 research scholars (Ph.D.) and M.Tech. students for their dissertation work. He is author/co-author of many research articles published in international refereed journals (including IEEE transactions) and international conferences. Mr. Sharma is a life member of I.E.T.E., India. Presently, he is coordinator of an International Conference to be held in January, 2013 at Malaviya National Institute of Technology, Jaipur, India

Shri Gopal Modani received B.E. degree in Electrical Engineering from University of Rajasthan (Malaviya Regional Engineering College, Jaipur), in 1970, the M.E. degree in Applied Electronics & Servo Mechanism from Indore Univ., SGSITS, Indore, in 1973, and the Ph.D. degree in Electronics Stream from University of Rajasthan (Malaviya Regional Engineering College, Jaipur), in 1987. He is Senior Professor with Department of Electronics & Communication Engineering from 1999 and working since last 35 years. His research interests include digital signal processing, electronic measurement techniques, Servo Mechanism and VLSI. He was deputed to Principal, Govt. Engineering College, Ajmer, India from 2004 to 2009 and member secretary, Technical Education Department, Govt. of Rajasthan, India. He was founder principal of Govt. Women Engineering College, Ajmer, India and Govt. Engineering College, Jhalawar, India. He was holding various important positions in the Technical Education Department, Govt. of Rajasthan, India. Presently, he is supervising more than 20 research scholars and dissertation work of M.Tech. students. He is author/and co-author of many research articles published in international refereed journals and international conferences.