IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 2, Ver. I (Mar. - Apr. 2015), PP 01-09
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

An Improved Speech De-noising Method based on Empirical Mode Decomposition

Mr. S. Nageswara Rao[1], Dr. K. Jaya Sankar[2], Dr. C.D. Naidu[3]
[1]*Assistant Professor, Department of ECE, Sri Venkateswara Engineering College, Suryapet*
[2]*Prof. & Head, Department of ECE, Vasavi College of Engineering, Hyderabad*
[3]*Professor, Department of ECE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad*

Abstract: *Generally, Speech enhancement aims to improve speech quality and intelligibility of a noise contaminated speech signal by using various signal processing approaches. Removal of a noise from a noisy speech is a common problem; already a vast research was carried out in earlier. However, due to the characteristics of various types of noises, the approaches proposed in earlier are not applicable for all types of noises. In addition, the earlier approaches didn't focus on the non-linear and non-stationary characteristics on noise environments. EMD is a filtering approach performs efficiently for non-stationary environments. This paper proposes a novel EMDF approach with the inspiration of thresholding to remove the noise from noisy speech sample. The proposed approach also developed a method to select the IMF index for separating the residual low-frequency noise components from the speech estimate, based on the IMF statistics. An experimental study was also done on various types of noise contaminated speech samples like babble noise, restaurant noise and car interior noise at various strengths.*

Keywords: *Speech enhancement, EMDF, IMF, noise estimation, SegSNR.*

I. Introduction

The enhancement of speech corrupted by noise is an important problemwith numerous applicationsranging from suppression of environmental noise for communication systems and hearing aids, enhancing the quality of old records, to preprocessing for speech recognition systems, cell phones and speech coding applications.Speech signals from the uncontrolled environment may contain degradation components along with the required speech components. Degradation components include back ground noise¸ reverberation and speech from other speakers. Therefore the degraded speech components need to be processed for the enhancement. Speech enhancement [1] algorithms improve the quality and intelligibility of speech by reducing or eliminating the noise component from the speech signals. Improving quality and intelligibility of speech signals reduce listener's exhaustion; improve the performance of hearing aids¸ speech coders and many other speech processing systems. In most speech enhancement algorithms it is assumed that an estimate of noise spectrum is available. Noise estimate is critical part and it is important for speech enhancement algorithms. Performance of speech enhancement algorithms depends on correct estimation of noise.

In earlier there are so many speech enhancement approaches were proposed to estimate and filter the noise from a noise contaminated speech signal. Voice activity detection (VAD) [2], [3] is a simple noise estimating approach, estimates the noise during the silent segments of speech. But the main problem, VAD is a conservative approach and it is less frequent to the noise updates and also not efficient for non-stationary environments. For highly non-stationary environments, the noise power has to track for speech segments also. Short time Fourier transform (STFT) domain is one of the frequency domain which gives the clear knowledge about the speech and noise during the noise estimation. Minimum statistics MS [4] and improved minima controlled recursive averaging (IMCRA) [5] are the two different approaches based on STFT, estimate the noise spectrum based onthe observation that the noisy signal power decays to values characteristic of the contaminating noise during speech pauses. However, the main problem, not able to track the noise power during speech segments, results poor estimation during long speech segments with few speech pauses. OMLSA is one of the speech enhancement approach proposed in [6], enhance the speech by estimating the noise with respect to spectral gain function, which minimizes the mean-square error of the log-spectral amplitude, is obtained as a weighted geometric mean of the hypothetical gain associated with the presence noise uncertainty. In [7], speech enhancement in car interior noise is achieved by using a speech analysis–synthesis approach, based on a harmonic noise model, as post processing after a traditional log-spectral amplitude speech estimation system. This system is sensitive to accurate pitch estimation and voiced/unvoiced speech frame classification.

Recently a new method for analyzing nonlinear and non-stationary data has been developed.The key part of the method is the empirical mode decomposition (EMD) [8], [9] methodwith which any complicated data set can be decomposed into a finite and oftensmall number of intrinsic mode functions.This decomposition method is adaptive, and, therefore, highly efficient. Sincethe decomposition is based on the local characteristic time scale of the data, it isapplicable to nonlinear and non-stationary processes.The EMD methods proposed in [8] and [9] are applicable for the fractional Gaussian noise only, but for a highly non-stationary noise environments like car noise, restaurant noise, babble noise etc., the earlier EMD based methods are not efficient.

This paper propose an efficient empirical mode decomposition filtering (EMDF) approach to estimate the noise based on the second order characteristics of Intrinsic mode functions IMFs for highly non-stationary environments. This approach decomposes the noise contaminated speech into individual IMFs and the filters the low frequency noise sample based on the selection of IMFs indices.

The rest of the paper is organized as follows; section II gives the basic details about the EMDF. Section III gives the details about the earlier EMD based approach. The complete illustration about the proposed approach is given in section IV. Section V gives the details about the performance evaluation of proposed approach and finally the conclusions are given in section VI.

II. Empirical Mode Decomposition

EMD is a non-linear decomposition technique for analysis and representation of non-stationary signals. The EMD method is a necessary step to reduce any given data into a collection of intrinsic mode functions (IMF) to which the Hilbert spectral analysis can be applied. EMD decomposes a signal in the time domain into adaptive basis functions which can be called as intrinsic mode functions (IMF). These data adaptive basis functions give physical meaning to the underlying process. The frequency information of the signal is embedded into IMFs.An IMF is defined as a function that satisfies the following requirements:

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

The procedure of extracting an IMF is called sifting. The sifting process is as follows:

1. Identify all the local extrema in the test data.
2. Connect all the local maxima by a cubic spline line as the upper envelope.
3. Repeat the procedure for the local minima to produce the lower envelope.

The upper and lower envelopes should cover all the data between them. Their mean is m_1. The difference between the data and m_1 is the first component h_1:

$$X(t) - m_1 = h_1 \quad (1)$$

After the first round of sifting, a crest may become a local maximum. New extrema generated in this way actually reveal the proper modes lost in the initial examination. In the subsequent sifting process, h_1 can only be treated as a proto-IMF. In the next step, it is treated as the data, then

$$h_1 - m_{11} = h_{11} \quad (2)$$

After repeated sifting up to k times, h_1 becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (3)$$

Then, it is designated as the first IMF component from the data:

$$c_1 = h_{1k} \quad (4)$$

At the end of the decomposition, the data s(t) will be represented as a sum of nIMF signals plus a residue signal,

$$s(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \quad (5)$$

The finally obtained signal s(t) represents the partially reconstructed signal. It is reconstructed by summing the obtained IMFs with the residue signal as represented above.

III. Emd-Thresholding Based Denoising [9]

In [9] a speech enhancement approach was proposed based on EMD-thresholding inspired with wavelet thresholding. This approach is considered only for speech signals contaminated with fractional Gaussian noise (fGN). As like wavelet thresholding, EMD-thresholding cannot be applied directly. Let consider a noisy speech signal$x(t)$,

$$x(t) = s(t) + \sigma n(t) \quad (6)$$

Where

$s(t)$Is a noiseless speech signal,$n(t)$ is Gaussian distributed noise and σ is the noise variance.

The fundamental aim of thresholding is to set to zero all the components that are lower than a threshold related to the noise level, the threshold can be evaluated based on the noise variance $T = \sigma C$, where, C=$\sqrt{2 \ln N}$ (for universal threshold) is a constant and during the reconstruction the signal utilizes high amplitudes only. There are two types of thresholding, soft thresholding and hard thresholding.

EMD can also be treated as a sub-band like filtering results in a number of uncorrelated IMFs. In this approach the threshold was applied directly to the IMFs to extract the low frequency noise components. This approach performs the checking of each and every IMF with the predefined threshold.by applying the threshold directly to EMD, it can be represented as

$$\tilde{c}^{(i)}(t) = \begin{cases} c^{(i)}(t) & |c^{(i)}(t)| > T \\ 0 & |c^{(i)}(t)| \le T \end{cases} \quad (7)$$

For hard thresholding

$$\tilde{c}^{(i)}(t) = \begin{cases} \operatorname{sgn}(c^{(i)}(t))\left(|c^{(i)}(t)| - T\right) & |c^{(i)}(t)| > T \\ 0 & |c^{(i)}(t)| \le T \end{cases} \quad (8)$$

For soft thresholding,Where, in both of the thresholding cases, $\tilde{c}^{(i)}(t)$ indicates the i_{th}threshold IMF. After performing the thresholding operation, the signal was reconstructed and a generalized reconstruction of the signal is represented as

$$\tilde{s}(t) = \sum_{k=M1}^{M2} \tilde{c}^{(i)}(t) + \sum_{k=M2+1}^{L} c^{(i)}(t) \quad (9)$$

Where the introduction of parameters M1 and M2 gives us flexibility on the exclusion of the noisy low-order IMFs and on the optional thresholding of the high-order ones, which in white Gaussian noise conditions contain little noise energy.

IV. Proposed Approach

This section gives the complete illustration about the proposed approach. The new EMDF system for the proposed approach is shown in fig.1.

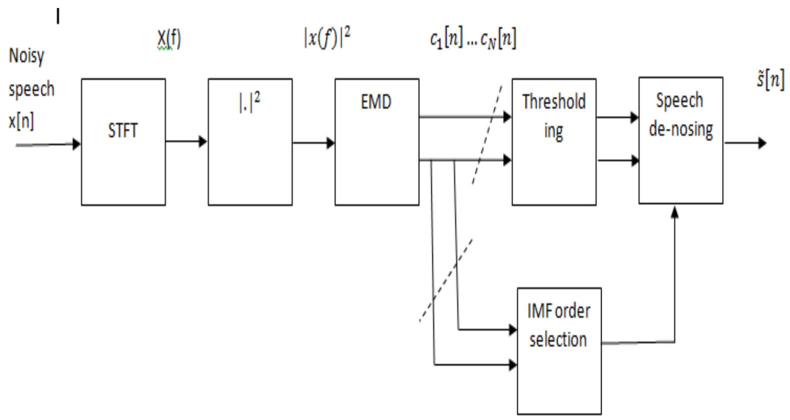


Fig.1. Block diagram of proposed EMDF system

Consider the noisy speech by

$$x(t) = s(t) + d(t) \quad (10)$$

Where$x(t)$ is the noisy speech signal, $s(t)$is the original noisefree speech, and$d(t)$ is the noise source which is assumed to be independent of the speech.In this approach STFT is applied at preprocessing stage to convert the time domain sample into frequency domain for frequency bin f and at time frame k. As shown in fig.1, the proposed approach performs the PSD estimation after STFT and then applies EMD operation to decompose the noisy speech into N IMFs $[c_1[n], c_2[n], \dots\dots\dots, c_N[n]]$.

The direct application of thresholding, either hard or soft, to the decomposition modes is in principle incorrect and can have shattering consequences for the continuity of the reconstructed signal.This is because the IMFs resemble an AM/FM modulated sinusoid with zero mean. As a result, it is guaranteed that, even in a noiseless case, in any interval$z_j^{(i)} = [z_j^{(i)}, z_{j+1}^{(i)}]$, the absolute amplitude of the i_{th} IMF, i=1,…,N, will drop below any nonzero threshold in the proximity of the zero-crossings $z_j^{(i)}$and $z_{j+1}^{(i)}$. It is possible to guess the interval noise dominant or signal dominant based on the product of obtained IMF and the residual at the stage. So according to this the modified threshold can be written as

$$\tilde{c}^{(i)}(z_j^{(i)}) = \begin{cases} c^{(i)}(z_j^{(i)}), & |c^{(i)}(r_j^{(i)})| > T \\ 0, & |c^{(i)}(r_j^{(i)})| \le T \end{cases} \quad (11)$$

For j=1,….,N.

At a time the obtained IMFs are processed to evaluate the order of IMFs. Let m be the order of IMFs and IMF variance is given as V[m] and can be obtained as

$$V[m] = \frac{1}{L}\sum_{n=1}^{L} c_m^2[n] \qquad (12)$$

Where $c_m[n]$is the mth IMF and then the partial reconstructed signal can be represented as

$$\tilde{s}[n] = \sum_{m=1}^{M} c_m(n) \qquad (13)$$

The order of IMF [10], M can be defined as an order by which there is sufficient IMFs to reconstruct the signal form IMFs. It varies from male to female and also from speech to speech.

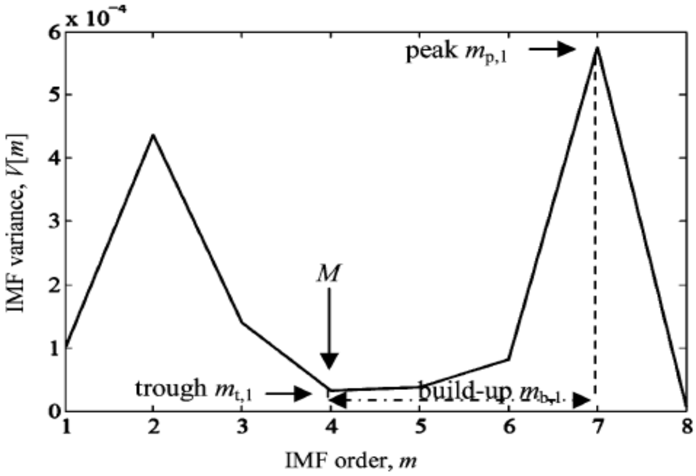


Fig.2. IMF variance plot of clean speech contaminated with car interior noiseat 0-dB SNR.

The order of the IMF M can be determined from the peak and troughs of IMF variance V as follows:

1. Initially compute the variance V of mthIMFc_m .
2. Identify the peaks m_p for m>4
3. If the peaks are not identified then order of IMF, M=N
   Else

Identify indices corresponding to troughs m_t
Compute IMF variance mb to peaks in m_p
Find the index i of the first occurrence for m_b
The order $M=m_{b,i}-m_{p,i}$,

4. Order of IMF=M.

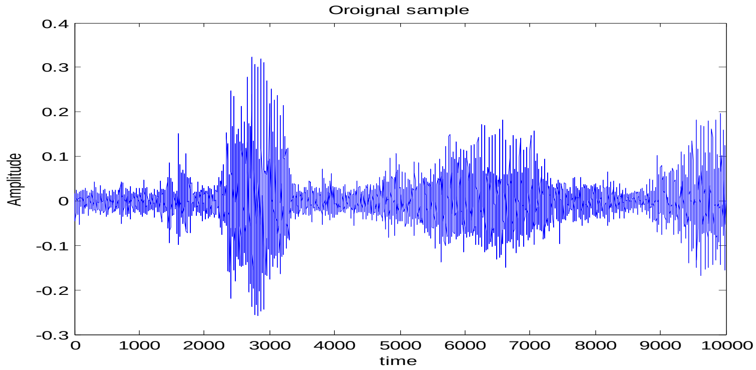
This method for the selection of M was used for filtering the residual low-frequency noise from the speech estimate to $\tilde{s}[n]$ in (13). The obtained speech estimate can be used to compare with earlier approaches.

V. Results

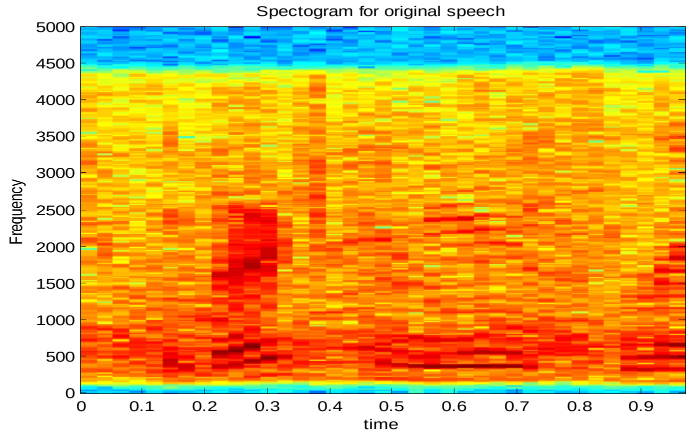
This section gives the complete details about the results evaluation. The proposed approach was tested over various types of noises. To verify the performance of proposed approach here a numerical parameter is evaluated, segSNR [10]. To test the proposed approach we have considered the three types of noisy speech signals. Those are babble noisy speech, restaurant noisy speech and car interior noisy speech. Now each source signal has 10,000 samples and having the sampling frequency of 16,000 Hz. The accuracy of the proposed approach is measured with respect to segSNR as

$$\text{segSNR} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\sum_{i=k}^{k+N-1}\left(\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right) \qquad (14)$$

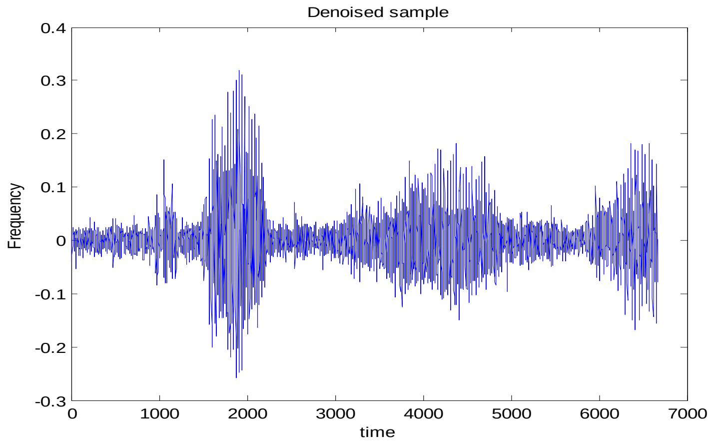
Where N and M are length of segment and number of segments respectively, $x(i)$and $y(i)$are the original and processed speech samples indexed by iand N is the total number of samples. The following figures give the description about the speech samples processed during the evaluation.



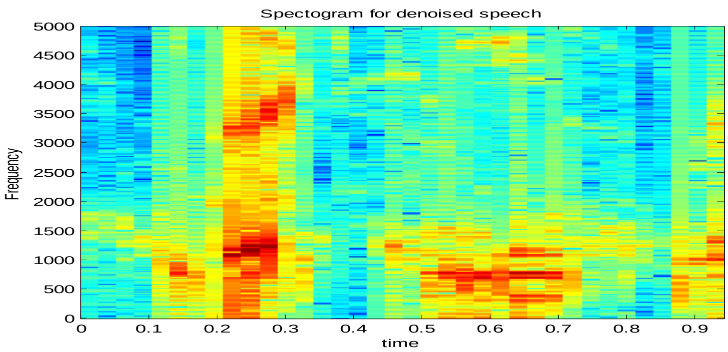
(a)



(b)



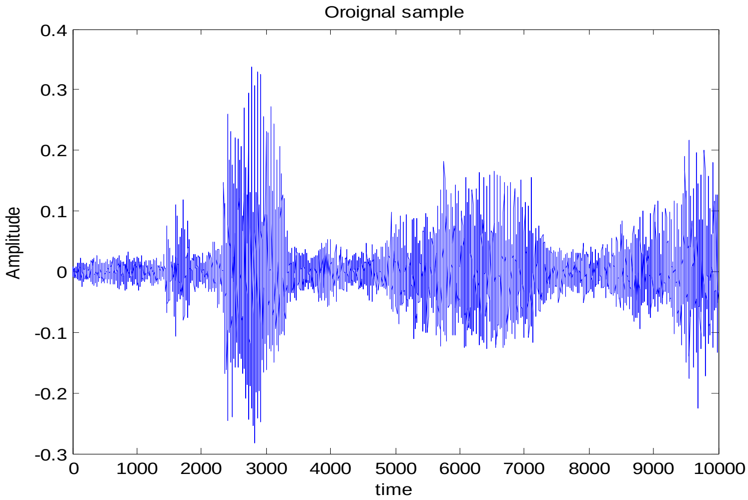
(c)



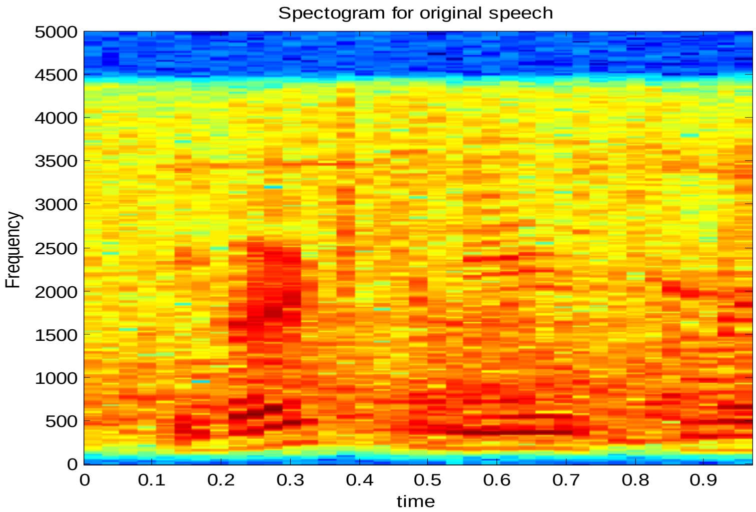
(d)

Fig.3. (a) clean speech signal with Babble noise signal at 5db, (b): Spectrogram, (c): Denoised sample,

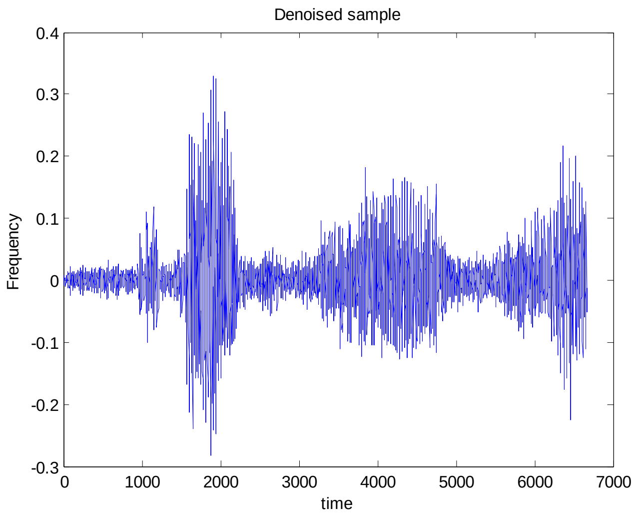
(d): Spectrogram of Denoised sample



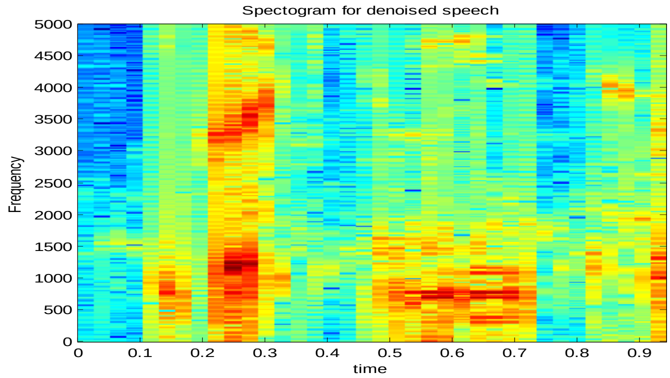
(a)



(b)

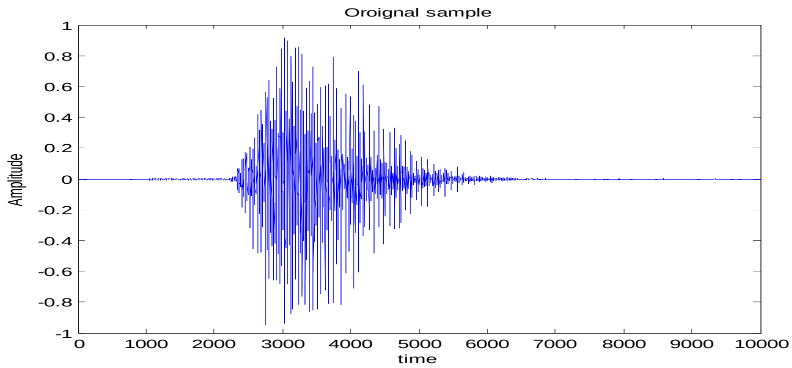


(c)

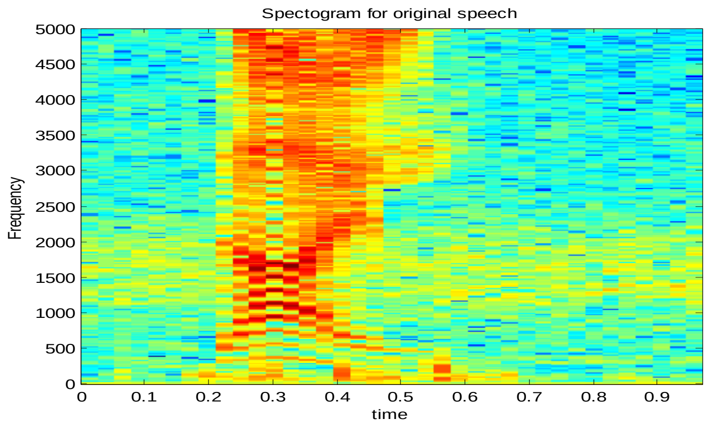


(d)

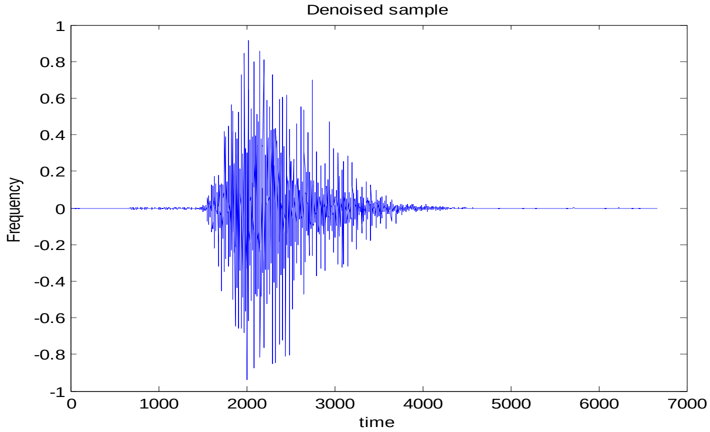
Fig.4. (a) clean speech signal with Restaurant noise at 5db, (b): Spectrogram, (c): Denoised sample, (d): Spectrogram of Denoised sample



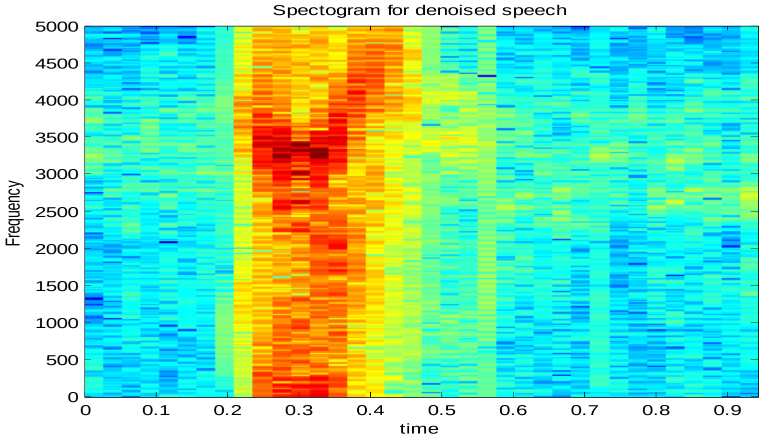
(a)



(b)



(c)



(d)

Fig.5. (a) clean speech signal with car interior noise at 5db, (b): Spectrogram, (c): Denoised sample, (d): Spectrogram of Denoised sample

The segSNR evaluated for the clean speech sample of male contaminated with babble noise, restaurant noise and car interior noise at various strengths and are shown in table I.

Table I. Seg SNR for various types of noises at various strengths for a male noisy speech signal

Input	Contaminated with	db	SegSNR
Clean Speech of Male	Babble Noise	5db	1.5737
	Babble Noise	10db	1.5619
	Babble Noise	15db	1.5421
	Restaurant Noise	5db	1.5282
	Restaurant Noise	10db	1.7769
	Restaurant Noise	15db	1.5318
	Car Interior noise	3db	4.1666
	Car Interior Noise	6db	4.2826

From the above table, it can be seen that the best overall improvements are obtainedunder car interior noisy conditions which is dominatedby low-frequency noise components. As shown in Table I, the proposed EMDF also achieves increased noise suppression and reduced speech distortion in babble noise conditions and restaurant noise conditions also.

VI. Conclusions

This paper proposed a new EMD-based filtering (EMDF) technique is described as a postprocessor for noisy speech which is enhanced using a thresholding based noise estimate. This proposed technique has been designed to be particularly effective in low-frequency noise environments. In EMDF, the speech is first decomposed into its IMFs using EMD. An adaptive method is developed to select the IMF index for separating the residual low-frequency noise components from the speech estimate, based on the IMF statistics. The performance of this technique was evaluated usingspeech contaminated with car interior noise, babble noise,and military vehicle noise conditions. When compared toaprevious approach, this method was shown to giveimproved performance at suppressing background noise underthe presented noisy conditions.

References

[1]. I. Cohen and B. Berdugo, “Speech enhancement for non-stationarynoise environments”, in Signal Processing. Amsterdam, the Netherlands:Elsevier, Nov. 2001, vol. 81, pp. 2403–2418.
[2]. Sohn J. and Kim N, “Statistical Model based voice activity detection”, IEEE Signal Processing Letters, Jan 1999, Volume:6 , Issue: 1) pp.1-3.
[3]. Shrinivasan K. and Gersho A. (1993), “Voice ActivityDetection for Cellular Network”, Prec. IEEE Speech CodingWorkshop pp. 85-86.
[4]. R. Martin, “Noise PSD estimation based on optimal smoothing andminimum statistics,” IEEE Trans. Speech Audio Process., vol. 9, no. 5,pp. 504–512, Jul. 2001.
[5]. I. Cohen, “Noise spectrum estimation in adverse environments: Improvedminima controlled recursive averaging,” IEEE Trans. SpeechAudio Process., vol. 11, no. 5, pp. 466–475, Sep. 2003.
[6]. Tien Dung Tran, “Speech enhancement using modified IMCRA and OMLSA methods”, Third International Conference on Communications and Electronics (ICCE), Aug. 2010.
[7]. Anuradha R. Fukane, Shashikant L. Sahare, “Noise estimation Algorithms for Speech Enhancement in highlynon-stationary Environments”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011ISSN (Online): 1694-0814.

[8]. P. Flandrinet al., “Detrending and denoising with empirical mode decompositions,”in Proc. Eur. Signal Process. Conf. (EUSIPCO), 2004,pp. 1581–1584.

[9]. YannisKopsinis, “Development of EMD-Based Denoising MethodsInspired by Wavelet Thresholding”, IEEE Transactions on Signal Processing, Vol. 57, No. 4, April 2009

[10]. NavinChatlani and John J. Soraghan, Senior Member, IEEE, “EMD-Based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement”, 1158 IEEE Transactions On Audio, Speech, And Language Processing, Vol. 20, No. 4, May 2012

[11]. G. Rillinget al., “On empirical mode decomposition and its algorithms,”in Proc. IEEE-EURASIP Workshop NSIP, Jun. 8–11, 2003.

[12]. P. Flandrin and G. Rilling, “Empirical mode decomposition as a filterbank,” IEEE Signal Process. Lett., vol. 11, no. 2, pp. 112–114, Feb.2004.

[13]. X. Zouet al., “Speech enhancement based on Hilbert-Huang transformtheory,” in Proc. IEEE CS 1st Int. Multi-Symp. Comput.Comput. Sci. (IMSCCS’06), 2006, pp. 208–213.

[14]. T. Hasan and M. K. Hasan, “Suppression of residual noise from speechsignals using empirical mode decomposition,” IEEE Signal Process. Lett., vol. 16, pp. 2–5, Jan. 2009.

Authors profile

S. Nageswara Rao is currently working as Assistant Professor in SVEC, Suryapet, Telangana. He received his M.Tech degree in Computer and Communication (ECE) from the Jawaharlal Nehru Technological University, Hyderabad in 2004. His current research interest is Digital Signal Processing Based Global Non-Parametric Frequency Domain Audio Testing.

Dr. K. Jaya Sankar obtained bachelor□ s degree in ECE from S.V. University, Tirupati in the year 1988. M.E., (Microwave and Radar Engineering) from Osmania University Hyderabad, in year 1994 and Ph.D. degree in 2004 from the same University. After around 4 years of industrial experience, he joined the Department of ECE at Vasavi College of Engineering in 1992. Presently he is Professor and Head of Department. He has around 20 Research Publications to his credit. His areas of interest include Electromagnetics, Antennas, Digital and Wireless Communication and Signal Processing.

Dr. C.D. Naidu received B.Tech degree in Electronics & Communication Engineering from JNTU Anantapur, M.Tech in Instrumentation and Control Systems from S.V. University, Tirupathi and Ph.D from JNTU Hyderabad. He is currently working as Principal of VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad.