5th International Conference on Control Engineering & Information Technology (CEIT-2022)
Proceedings Book Series –PBS- Vol 5, pp.30-35

# Speech enhancement in EMD domain using spectral subtraction and Wiener filter

Kais KHALDI[*#1], Haifa TOUATI[$2]

[*]*Unité Signaux et Sytèmes, ENIT, ElManar University*
*BP 37, Le Belevedère, 1002 Tunis, Tunisia*
[#]*College of Science and Arts-Tabarjal, Al Jouf University*
*P.O.Box 2014, Al-Jouf , Skaka,42421, KSA*
[$]*ISSAT, University of Gafsa,*
*BP 116, Sidi Ahmed Zarrouk , 2112 Gafsa, Tunisia*
[1]kais.khaldi@gmail.com
[2]haifatwati1991@gmail.com

*Abstract*— **This paper proposes a technique in Empirical Mode Decomposition (EMD) domain to enhance the signal. The noisy signal is decomposed, by EMD, into approximation and detail which are filtered separately using spectral subtraction and Wiener filter. Therefore, the main idea of the proposed approach is to filter the shorter scale IMF (detail) by Wiener filter, which are noise dominated, and filter the approximation using spectral subtraction technique. In fact, the filtering of the approximation by the same filter (Wiener) will introduce signal distortion rather than a noise reduction. Thus, the performance of this method is to construct linearly the original signal without loss of the useful information. The study is limited to signals corrupted by additive white Gaussian noise.**

*Keywords*— **Empirical Mode Decomposition, Wiener filter, Spectral Subtraction , Speech enhancement, detail, approximation.**

## I. Introduction

Speech signal noise reduction is a well known problem in signal processing. Particularly, in the case of additive white Gaussian noise a number of filtering methods has been proposed[1]-[2]. However, these methods are not effective when the noise estimation is not possible. To overcome these difficulties, nonlinear methods have been proposed and especially those based on Wavelets thresholding [2]-[3]. The idea of wavelet thresholding relies on the assumption that signal magnitudes dominate the magnitudes of the noise in a wavelet representation, so that wavelet coefficients can be set to zero if their magnitudes are less than a pre-determined threshold [3]. A limit of the wavelet approach is that the basis functions are fixed, and thus do not necessarily match all real signals.

Recently, a new temporal signal decomposition method, called Empirical Mode Decomposition (EMD), has been introduced by Huang et al. [4] for analysing data from non stationary and nonlinear processes. The major advantage of the EMD is that the basis functions are derived from the signal itself. Hence, the analysis is adaptive in contrast to the traditional methods where the basis functions are fixed. In our previous works [5]-[6], the denoising method is based on the filtering of all IMFs extracted from the noisy signal by the same filter. However, the longer scale IMFs (low- and medium-frequency components) corresponding to the most important structures of the signal is signal dominated. Therefore, filtering of these IMFs will introduce signal distortion rather than a noise reduction [7]. The basic idea of the proposed method is to filter the shorter scale IMF (detail) by Wiener filter, which are noise dominated, and filter the approximation using spectral subtraction technique. In fact the filtering of all IMFs by the Wiener filter generates a distortion of the signal, i.e. the filtering eliminates even the useful information. While the filtering of all IMFs by the spectral subtraction filter does not make it possible to effectively eliminate all the noise.

The paper is organized as follows. Section II explains the basics of the EMD and Section III exposed the proposed method. Results are presented in Section IV, and conclusions are drawn in Section V.

## II. EMD Basics

The EMD decomposes a signal f(t) into a series of IMFs through an iterative process called sifting; each one, with distinct time scale [8]. The decomposition is based on the local time scale of f(t) and yields adaptive basis functions. The EMD can be seen as a type of wavelet decomposition whose subbands are built up as needed to separate the different components of f(t). Each IMF replaces the signals detail, at a certain scale or frequency band [9]. The EMD picks out the highest-frequency oscillation that remains in f(t). By definition, an IMF satisfies two conditions:

- the number of extrema and the number of zero crossings may differ by no more than one;
- the average value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

ISSN: 2961-6611

Thus, locally, each IMF contains lower-frequency oscillations than the just-extracted one. To be successfully decomposed into IMFs, f(t) must have at least two extrema; one mini-mum and one maximum. The sifting involves the following steps:

Step 1: fix the threshold $\epsilon$ and set j ← 1 ( jth IMF);
Step 2: $r_{j-1}$(t) ← f(t) (residual);
Step 3: extract the jth IMF:

(a) $h_{j,i-1}$(t) ← $r_{j-1}$(t), i ← 1 (i number of sifts).
(b) extract local maxima/minima of $h_{j,i-1}$(t).
(c) compute upper and lower envelopes $U_{j,i-1}$(t) and $L_{j,i-1}$(t) by interpolating, using cubic spline, respectively, local maxima and minima of $h_{j,i-1}$(t),

(d) compute the mean of the envelopes :

$$\mu_{j,i-1}(t) = \left( U_{j,i-1}(t) + L_{j,i-1}(t)\right)/2$$

(e) update: $h_{j,i}$(t) ←$h_{j,i-1}$(t) - $\mu_{j,i-1}$(t), i ← i + 1,
(f) calculate the stopping criterion :

$$SD(i) = \sum_{t=0}^{T} \frac{\left|h_{j,i-1}(t) - h_j(t)\right|^2}{(h_{j,i-1}(t))^2}$$

(g)repeat steps (b)-(f) until SD(i)< $\epsilon$ and then put $IMF_j(t)$←$h_{j,i}$(t) ( jth IMF).
Step 4: update residual : $r_j$(t):= $r_{j-1}$(t)- $IMF_j(t)$;
Step 5: Repeat step 3 with j := j+1 until the number of extrema in $r_j$(t) is ≤ 2;
where T is f(t) time duration. The sifting is repeated several times (i) in order to get h true IMF that fulfils the two first conditions. The result of the sifting is that f(t) will be decomposed into a sum of $C$ IMFs and a residual $r_c$(t) such that

$$f(t) = \sum_{j=1}^{C} IMF_j(t) + r_c(t) \qquad (1)$$

C value is determined automatically using SD (Step 3(f)). The sifting has two effects: (a) it eliminates riding waves and (b) to smoothen uneven amplitudes. To guarantee that IMF components retain enough physical sense of both amplitude and frequency modulation, we have to determine SD value for the sifting. This is accomplished by limiting the size of the standard deviation SD computed from the two consecutive sifting results. Usually, SD (or ) is set between 0.2 to 0.3 [8].

## III. THE PROPOSED METHOD

The proposed approach denoising is illustrated by Fig. 1. The EMD breaks in first step the noisy signal into low frequency components known as the approximation and high frequency components known as the detail (IMF).

The approximation is filtered using the spectral subtraction which suppresses the stationary noise components. The detail is filtered using the Wiener filter which is employed to reduce the real background noise. Then, the denoised signal reconstructed by sum the filtered detail and approximation

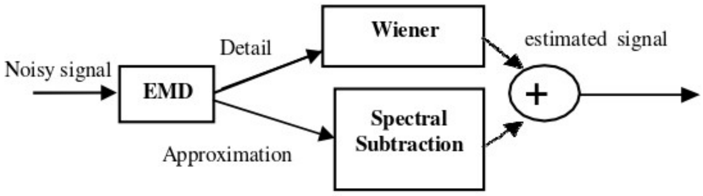

Fig. 1. The proposed approach denoising scheme

### A. *Wiener filter*

Let a clean speech signal s(t) be corrupted by an additive white Gaussian noise w(t) as follows:

$$x(t) = s(t) + w(t) \qquad (2)$$

Wiener filter is simple, easy to implement and to design[1]. Also, it's an optimal filter that minimizes the Mean Squared Error (MSE) criterion [10]. In (2), The filter defined by:

$$S(\omega) = H(\omega)\, X(\omega) \qquad (3)$$

where ω denote the frequency index, S(ω),X(ω), and H(ω) are the discrete Fourier transform of the clean speech, the transform of noisy speech and transfer function of Wiener filter, respectively. The MSE can be defined as follows. Also, The Wiener filter can be derived by:

$$H(\omega) = \frac{P_{ss}(\omega)}{P_{ss}(\omega) + P_{ww}(\omega)} \qquad (4)$$

where $P_{ss}(\omega)$and $P_{ww}(\omega)$ denote power spectrum of speech, s(t) and that for noise, w(t), respectively.

In the case of (3) and (4), the enhanced speech is estimated in the frequency domain by:

$$H(\omega) = \frac{P_{ss}(\omega)}{P_{ss}(\omega) + P_{ww}(\omega)} X(\omega) \qquad (5)$$

### *B.* Spectral Subtraction

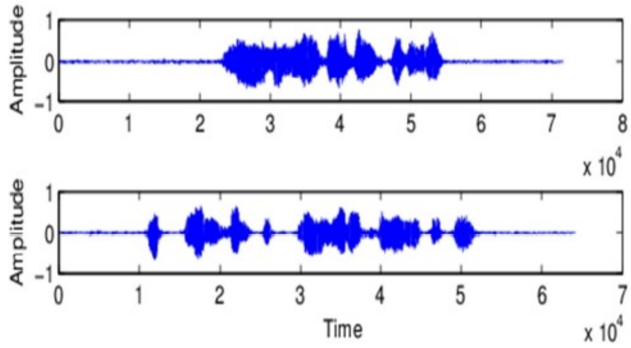

Fig. 3. Denoising results of signals "speech1" and "speech2" by the proposed method

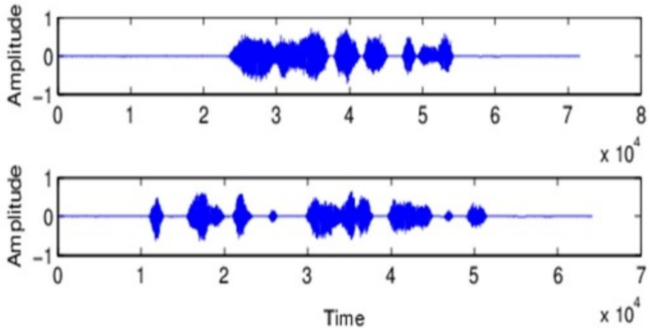

Fig. 4. Denoising results of signals "speech1" and "speech2" by the Wiener filter

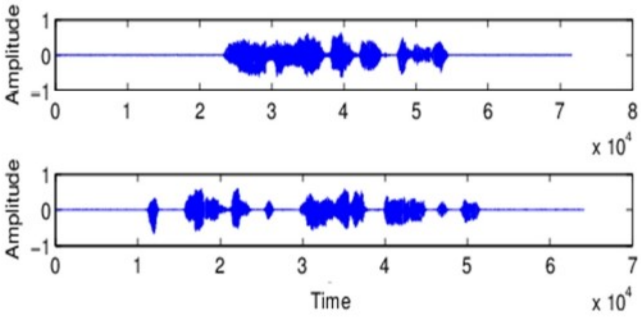

Fig. 5. Denoising results of signals "speech1" and "speech2" by the Spectral subtraction

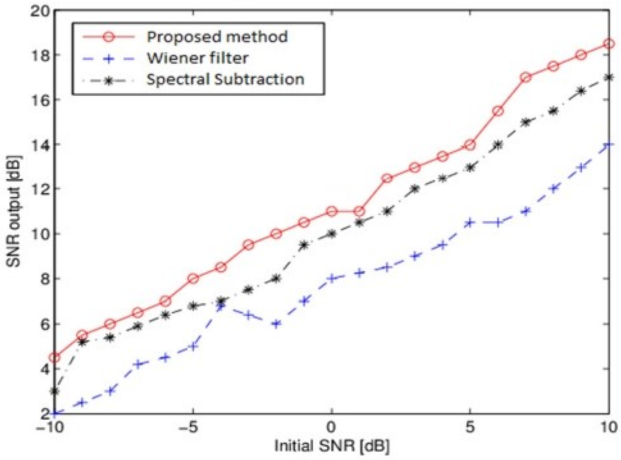

Fig. 6. Final SNR values obtained from different initial noise levels of signal "speech1". The results are averages over 100 instances of the noisy signals. They are reported for proposed method, Wiener filter and spectral subtraction.

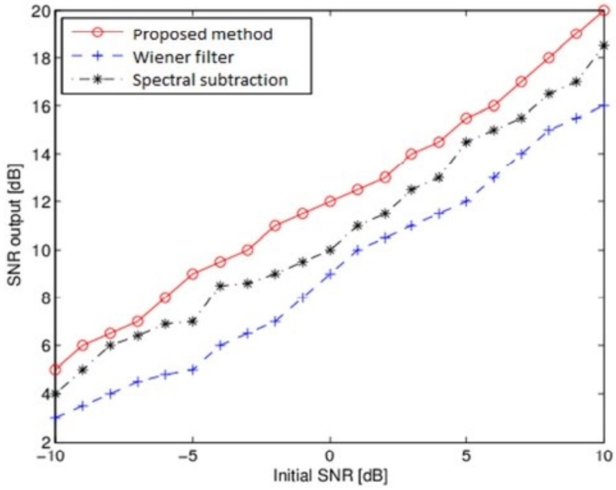

Fig. 7. Final SNR values obtained from different initial noise levels of signal "speech2". The results are averages over 100 instances of the noisy signals. They are reported for proposed method, Wiener filter and spectral subtraction.

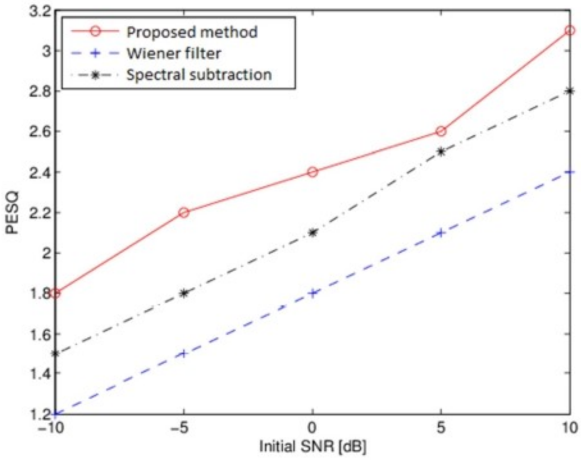

Fig. 8. PESQ values obtained from different initial noise levels of signal "speech1". The results are averages over 100 instances of the noisy signals. They are reported for proposed method, Wiener filter and spectral subtraction.

Spectral Subtraction method is one of the important speech enhancement methods, due to its relative simplicity and ease of implementation. Let x (t) a noisy speech signal:

$$x(t) = s(t) + w(t) \tag{6}$$

where s(t) is the clean signal and w(t) is the white Gaussian noise. The basic principle of spectral subtraction is to restore the power spectrum on an observed signal corrupted with additive noise, through subtraction of an estimate of noise spectrum from the noisy signal spectrum. The noisy spectrum is estimated from the periods of silence when the signal is absent and only the noise is present [11]. The equation describing Spectral Subtraction can be defined by:

$$\left|\hat{S}(\omega)\right|^2 = |X(\omega)|^2 - \hat{P}_{w(}\omega) \approx |X(\omega)|^2 \left[1 + \frac{1}{SNR}\right]^{-1} \tag{7}$$

where ω denote the frequency index, S(ω) and X(ω) and $\hat{P}_{w(}\omega)$ are the discrete Fourier transform of the clean speech, the transform of noisy speech, and the estimate of the power spectrum of noise, respectively. Finally in this case, the SNR defined :

$$SNR = \frac{|X(\omega)|^2}{\hat{P}_{W(\omega)}} \tag{8}$$

## C. *Noise estimation*

Extensive simulations have shown that when a speech signal with a silence sequence is decomposed by EMD, its first IMF corresponds to that silence sequence. Thus, the first IMF can be used to correctly estimate the noise level [5].
Generally, speech noise estimation is performed using the Boll's method. Accordingly, the silence periods of the signal are detected, and then power spectra noise estimation is performed by considering the average of the power spectra of the noisy signal on the M first temporal frames which are considered as being moments of silence, following the relation

$$\left|\widehat{W}(fe, m)\right|^2 = \frac{1}{M}\sum_{i=0}^{M-1}|W(fe, i)|^2 \tag{9}$$

where |W(fe, i)| is power spectra value at the discrete frequency fe of frame i. This method gives a correct estimation of the noise.

## IV. RESULTS

The proposed approach is applied to two clean speech signals "speech1" and "speech2" corrupted by additive white Gaussian noise with input SNR values ranging from -10 dB to 10dB. The original signals and the noisy versions corresponding to input SNR=-2 dB are shown in Fig .2.

The results are compared to the Wiener filter denoising technique and to the spectral subtraction technique. Where the Wiener filter denoising technique consists on filtering the noisy speech by Wiener filter, and the spectral subtraction technique consists on filtering the noisy speech by Spectral subtraction. The output SNR and Perceptual Evaluation of Speech Quality (PESQ) [12] are used as an objective measure to evaluate the denoising methods. More precisely, the PESQ criterion measures the perceptual quality of speech signal.

Fig.3,4 and 5 shows the denoising results obtained by the proposed method, the Wiener filter and the spectral subtraction technique. From these figures, one can conclude that the proposed approach performs better (noise reduction) than Wiener filter and Spectral subtraction technique compared to the original signals (Fig.2).

This fact is confirmed by the results shown in Fig. 6 and 7, where more SNR gain is obtained by the proposed approach compared to the Wiener and spectral subtraction. For each input SNR value, 100 independent noise simulations are generated and the average of output SNR and the PESQ values are calculated. One may note that the proposed approach provides an improvement about 1 dB compared to the standard Wiener filter and spectral subtraction technique for the noisy versions of all signals "speech1" and "speech2".

The obtained results also show that it is more efficient to apply the Wiener filter for detail signal and the spectral subtraction for approximation signal than to the signal itself. These results are also demonstrated by Fig.8 and Fig.9, where the PESQ values obtained from the proposed approach are better than those corresponding to Wiener filter and spectral subtraction technique.

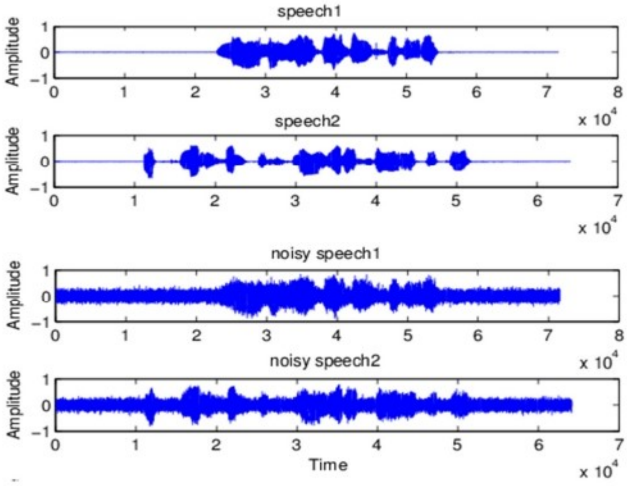

Fig. 2. The original and noisy version (input SNR=-2db) of signals "speech1" and "speech2"

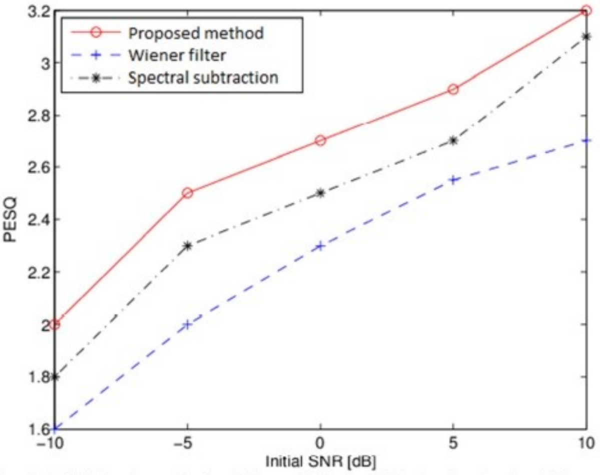

Fig. 9. PESQ values obtained from different initial noise levels of signal "speech2". The results are averages over 100 instances of the noisy signals. They are reported for proposed method, Wiener filter and spectral subtraction.

## V. Conclusions

In this paper, a new speech enhancement method is presented. To lower the noise level, two effective and powerful methods, Wiener filtering for the detail and spectral subtraction filtering for the approximation, are combined. Obtained results for denoising speech signals with different SNR values ranging from −10 dB to 10 dB show that the SNR improvement achieved by the proposed method is higher than those achieved by the Wiener filter and the spectral subtraction method. In addition, the PESQ criterion confirms that the proposed method offers a much better listening quality than the other methods. To confirm the obtained results and the effectiveness of the EMD-denoising approach, the scheme must be evaluated with large class of speech signals and in different experimental conditions such as sampling rates, sample sizes, multiplicative noise, or the type of noise.

## References

[1] J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed, Prentice-Hall , 1996.

[2] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform.Theory*, vol. 41, pp. 613–627, May. 1995.

[3] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrica*, vol. 81, pp.425–455, Sept.1994.

[4] N.E. Huang et al.,"The empirical mode decomposition and Hilbert spectrum for non linear and non-stationary time series analysis," *Proc. Royal Society*, vol. 454, pp. 903–995, Mar. 1998.

[5] K. Khaldi, A.O. Boudraa, A. Bouchikhi, and M. Turki, "Speech Enhancement via EMD", *EURASIP Journal Advances in Signal Processing,* vol. 2008, Article ID 873204, 8 pages, 2008.

[6] K. Khaldi, A.O. Boudraa and M. Turki, "Voiced/unvoiced speech classification-based adaptive filtering of decomposed empirical modes for speech enhancement", *IET Signal Processing*, Vol. 10, pp. 69 – 80, Jan. 2016.

[7] K. Khaldi, M. Turki and A.O. Boudraa, "Voiced speech enhancement based on adaptive filtering of selected Intrinsic Mode Functions", *Advances in Adaptive Data Analysis (AADA),*vol. 2, pp. 65-80, Jan. 2010.

[8] N. E. Huang, Z. Shen, S. R. Long, et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.

[9] P. Flandrin, G. Rilling, and P. Gonc¸alves,` "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, pp. 112–114, 2004.

[10] K. Funaki, "Speech enhancement based on iterative Wiener filter using complex speech analysis", *IEEE Signal Processing,* p.4, japan,2008.

[11] M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd El-samie, "Speech enhancement using an adaptive Wiener filtering approach", *Progress In Electromagnetics Research M*, Vol..4, pp. 167–184, 2008.

[12] *ITU-T P.835. Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.* ITU T RecommendationP.835,2003.