Speech Signal Noise Reduction by EMD

Conference Paper · April 2008

DOI: 10.1109/ISCCSP.2008.4537399 · Source: IEEE Xplore

Kais Khaldi
National Engineering School of Tunis

Abdel-O. Boudraa
French Naval Academy | Arts & Metiers Institute of Technology

Bouchikhi Abdelkhalek

Turki Hadj Alouane Monia
King Khalid University

SPEECH SIGNAL NOISE REDUCTION BY EMD

Kais KHALDI[*]†, *Abdel-Ouahab BOUDRAA*[†‡], *Abdelkhalek BOUCHIKHI*[†‡],
Monia TURKI-HADJ ALOUANE[*] *and El-Hadji Samba DIOP*[†‡]

[*]Unité Signaux et Systèmes, ENIT
BP 37, Le Belvedre 1002 Tunis, Tunisia.
[†]IRENav, Ecole Navale/[‡] $E^3I^2(EA3876)$, ENSIETA
Groupe ASM, Lanvéoc Poulmic, BP600, 29240 Brest−Armées, France.
(bouchikhi,boudra,diop)@ecole-navale.fr, khaldikais@yahoo.fr, m.turki@enit.rnu.tn

ABSTRACT

In this paper, a speech signal noise reduction based on a multiresolution approach referred to as Empirical Mode Decomposition (EMD) [1] is introduced. The proposed speech denoising method is a fully data-driven approach. Noisy signal is decomposed adaptively into oscillatory components called Intrinsic Mode Functions (IMFs), using a temporal decomposition called sifting process. The basic principle of the method is to reconstruct the signal with IMFs previously thresholded using a shrinkage function. The denoising method is applied to speech with different noise levels and the results are compared to wavelet shrinkage. The study is limited to signals corrupted by additive white Gaussian noise.

1. 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 [2]-[3]. Linear methods such as the Wiener filtering [2] are largely used because linear filters are easy to implement and to design. 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 [4]-[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 temporel signal decomposition method, called Empirical Mode Decomposition (EMD), has been introduced by Huang et al. [1] for analyzing data from nonstationary 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 this paper, a denoising scheme using EMD is proposed. The EMD is based on the sequential extraction of the energy associated with various intrinsic time scales of the signal, starting from finer temporal scales (high frequency modes) to coarser ones (low frequency modes). The total sum of the IMFs matches the signal very well and therefore ensures completeness. The basic idea of the proposed method is to pre-process each IMF using a thresholding, as in Wavelets analysis, or filtering before complete signal reconstruction. This method is applied to speech signals corrupted with different noise levels, and the results are compared to wavelets approach.

2. EMD ALGORITHM

The EMD decomposes a given signal $x(t)$ into a series of IMFs through an iterative process called *sifting*; each one with a distinct time scale [1]. The decomposition is based on the local time scale of $x(t)$, and yields adaptive basis functions. The EMD can be seen as a type of wavelet decomposition whose sub-bands are built up as needful to separate the different components of $x(t)$. Each IMF replaces the signal details, at a certain scale or frequency band [5]. The EMD picks out the highest frequency oscillation that remains in $x(t)$. By definition, an IMF satisfies two conditions :

1. the number of extrema and the number of zeros crossings may differ by no more than one.
2. the average value of the envelope defined by the local maxima, and the envelope defined by the local minima, is zero.

Thus, locally, each IMF contains lower frequency oscillations than the just extracted one. The EMD does not use a predetermined filter or a wavelet function, and is a fully data-driven method [1]. To be successfully decomposed into IMFs, $x(t)$ must have at least two extrema, one minimum and one maximum. The sifting involves the following steps :

978-1-4244-1688-2/08/$25.00

Step 1: Fix the threshold ϵ and set $j \leftarrow 1$ (j^{th}IMF)
Step 2: $r_{j-1}(t) \leftarrow x(t)$ (residual)
Step 3: Extract the j^{th}IMF :
(**a**) : $h_{j,i-1}(t) \leftarrow r_{j-1}(t)$,$i \leftarrow 1$ (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 : $\mu_{j,i-1}(t)$ =($U_{j,i-1}(t)$ + $L_{j,i-1}(t)$)/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 :

$$\text{SD(i)} = \sum_{t=1}^{T} \frac{|hj,i-1(t) - hj,i(t)|^2}{(hj,i-1(t))^2}$$

(**g**) : Repeat Steps (b)-(f) until SD(i)< ϵ and then put $\text{IMF}_j(t) \leftarrow h_{j,i}(t)$ (j^{th}IMF)
Step 4: Update residual : $r_j(t) := r_{j-1}(t) - \text{IMF}_j(t)$.
Step 5: Repeat Step 3 with $j := j+1$ until the number of extrema in $r_j(t)$ is ≤ 2.

Where T is $x(t)$ time duration. The sifting is repeated several times (i), in order to get h true IMF that fulfills the conditions (1) and (2). The sifting result is that $x(t)$ will be decomposed into a sum of C IMFs and a residual $r_C(t)$ as follows:

$$x(t) = \sum_{j=1}^{C} \text{IMF}_j(t) + r_C(t) \tag{1}$$

C value is determined automatically using SD (Step 3(f)). The sifting has two effects : (a) it eliminates riding waves, and (b) it smoothes 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, ϵ is set between 0.2 and 0.3 [1].

3. DENOISING PRINCIPLE

Let $x(t)$ be a clean signal corrupted by an additive white Gaussian noise $b(t)$ as follows:

$$y(t) = x(t) + b(t) \tag{2}$$

By EMD the noisy signal is expressed as:

$$y(t) = \sum_{j=1}^{C} \text{IMF}_j(t) + r_C(t) \tag{3}$$

where IMF_j is a noisy version of the data f_j:

$$\text{IMF}_j(t) = f_j(t) + b_j(t) \tag{4}$$

An estimation $\tilde{f}_j(t)$ of $f_j(t)$ based on the noisy observation $\text{IMF}_j(t)$ is given by

$$\tilde{f}_j(t) = \Gamma[\text{IMF}_j(t); \tau_j] \tag{5}$$

where $\Gamma[\text{IMF}_j(t); \tau_j]$ is a thresholding function (shrinkage), and τ_j is the threshold parameter, applied to signal IMF_j [6]-[7]. Finally, the denoised or reconstructed signal, $\tilde{x}(t)$, is given by :

$$\tilde{x}(t) = \sum_{j=1}^{C} \tilde{f}_j(t) + r_C(t) \tag{6}$$

The threshold parameter is given by [4],[6]-[7], [8]-[9]:

$$\tau_j = \sqrt{2\log(T)}\tilde{\sigma}_j \tag{7}$$

where $\tilde{\sigma}_j$ is the estimated noise level of IMF_j (scale level). According to [5] or [10], the noise level $\tilde{\sigma_j}$ of the IMFs can be estimated as follows

$$\tilde{\sigma_j} = \frac{\tilde{\sigma_1}}{\sqrt{2}^{j-1}} \tag{8}$$

where $\tilde{\sigma_1}$ is given by [6]-[7], [11]

$$\tilde{\sigma_1} = 1.4826 \times \text{Median}\left\{|\text{IMF}_1(t) - \text{Median}\left\{\text{IMF}_1(t)\right\}|\right\} \tag{9}$$

There are different non-linear shrinkage functions [12]. In the present work, we use the hard shrinkage which has given interesting denoising results for speech enhancement:

$$\tilde{f}_j = \begin{cases} \text{IMF}_j(t), & \text{if } |\text{IMF}_j(t)| > \tau_j \\ 0, & \text{if } |\text{IMF}_j(t)| \leq \tau_j \end{cases} \tag{10}$$

4. RESULTS

The proposed noise reduction method is tested on speech signals corrupted by additive white Gaussien noise whose level is fixed through the input Signal to Noise Ratio (SNR),

$$\text{SNR}_{\text{in}} = 10\log_{10}\frac{\sum_{t=1}^{T}(x(t))^2}{\sum_{t=1}^{T}(y(t)-x(t))^2} \tag{11}$$

where $x(t)$ and $y(t)$ are respectively the clean and the noisy signals respectively. The results obtained by the proposed method are compared to the wavelet approach (Daubechies 4, Symmlet 4 and Haar). As an objective criterion to evaluate the performance of the denoising method, we use the output SNR:

$$\text{SNR}_{\text{out}} = 10\log_{10}\frac{\sum_{t=1}^{T}(\tilde{x}(t))^2}{\sum_{t=1}^{T}(x(t)-\tilde{x}(t))^2} \tag{12}$$

The EMD-Shrinkage method and the wavelets method are applied to two clean speech signals "a" and "b" presented in figure 1. These signals are corrupted by a noise with SNR values ranging from -2dB to 3dB. Noisy version of the original signals corresponding to SNR=-1dB are shown in figure

2. For each SNR value, 100 independent noise sequences are generated and averaged values of the SNR_{out} are calculated. Denoising results of the EMD-Shrinkage (hard thresholding) and the wavelets method (Daubechies 4) are shown in figures 3 and 4 for an SNR =-1dB. A careful examination of the signals shown in figures 1, 3 and 4 shows that the EMD-Shrinkage performs better than the wavelet method in terms of noise reduction. Furthermore, signals structures or features are globally better preserved with the EMD-Shrinkage. Figures 5 and 6 show the variations of the output SNR versus the input SNR for the signals ”a” and ”b” respectively. The results obtained from EMD method and three wavelets methods are superposed. These figures show that the improvement in SNR provided by the EMD-Shrinkage varies from 8dB to 10.5dB. It is higher than the SNR improvement achieved by the three wavelets methods. When listening to the enhanced speech, the EMD-Shrinkage produce lower residual noise and noticeably less speech distortion for all the signals compared to the wavelets method.

5. CONCLUSION

This paper presents a new speech denoising method. The proposed scheme is based on the EMD and, consequently, is a simple and fully data-driven method. The method do not use any pre- or post-processing and do not require any user parameters setting (except the threshold ϵ). Obtained results for clean speech signals corrupted with additive Gaussian noise with different SNR values ranging from -2dB to 3dB show that the proposed EMD-denoising method, associated with the shrinkage strategy, performs better than the wavelets approach. These results show that the EMD-denoising method is effective for noise removal and confirm our findings presented in [6]-[7]. The EMD-Shrinkage is very attractive, especially when the noise estimation is not easy. The obtained results also show that it is more efficient to apply the thresholding to the different components (IMFs) of the signal than to the signal itself. 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.

6. REFERENCES

[1] N.E. Huang et al., “The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. Royal Society*, vol. 454, no. 1971, pp. 903–995, 1998.

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

[3] D.L. Donoho, “De-noising by soft-thresholding,” *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

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

[5] G. Rilling P. Flandrin and P. Goncalves, “Empirical mode decomposition as a filter bank,” *IEEE Sig. Proc. Lett.*, vol. 11, no. 2, pp. 112–114, 2004.

[6] A.O. Boudraa, J.C. Cexus, and Z. Saidi, “EMD-based signal noise reduction,” *Int. J. Sig. Process.*, vol. 1, no. 1, pp. 33–37, 2004, ISSN: 1304-4494.

[7] A.O. Boudraa and J.C. Cexus, “Denoising via empirical mode decomposition,” in *Proc. IEEE ISCCSP*, Marrakech, Morocco, 2006.

[8] D.L. Donoho and I.M. Johnstone, “Adapting to unknow smoothness via wavelet shrinkage,” *J. Am. Stat. Assoc.*, vol. 90, pp. 1200–1424, 1995.

[9] G. Kerkyacharian D.L. Donoho, I.M. Johnstone and D. Picard, “Wavelet shrinkage: Asymptopia with discussion,” *Proc. Royal Stat. Soc., Series B,*, vol. 57, pp. 301–396, 1995.

[10] G. Steidl, J. Weickert, T. Brox, Pavel Mrazek, and M. Welk, “On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs,” 2004.

[11] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, vol. 1, Cambridge Unive. Press, 1992.

[12] S. Mallat, *Une exploration des signaux en ondelettes*, Ellipses, 2000.

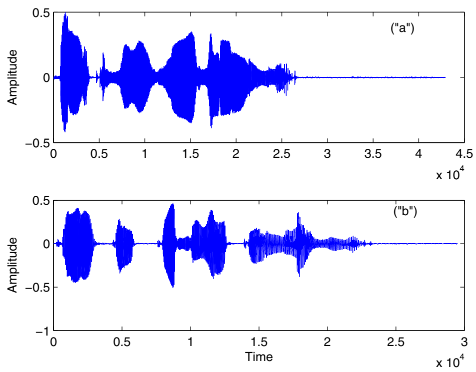

Fig. 1. The originalsignals ”a” and ”b”.

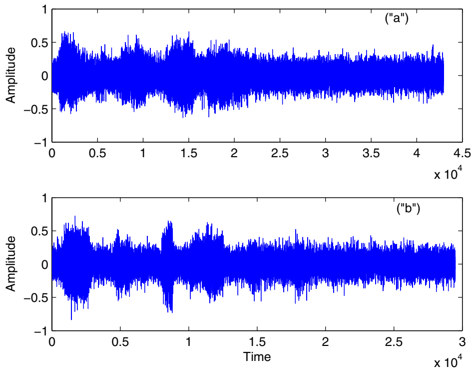

Fig. 2. The noisy version of signals ”a” and ”b” (SNR =-1dB).

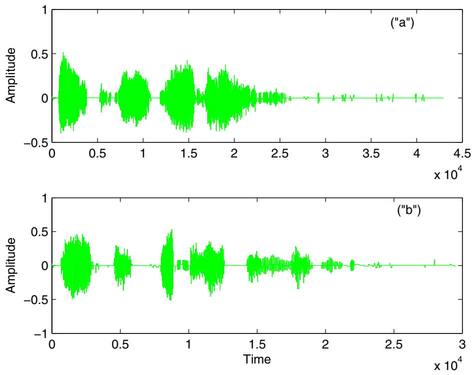

Fig. 3. Denoising results of signals ”a” and ”b” by the EMD-Shrinkage (SNR =-1dB).

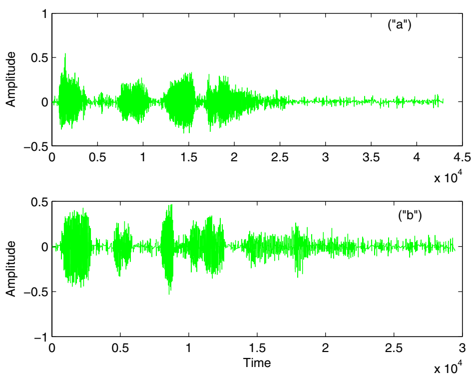

Fig. 4. Denoising results of signals ”a” and ”b” by the wavelets approach: Daubechies 4 (SNR=-1dB).

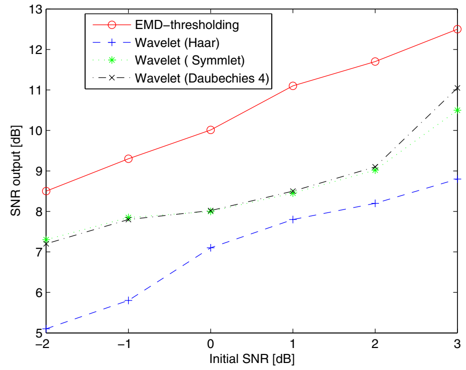

Fig. 5. Output SNR values obtained for different initial noise levels of signal ”a”. The results are the average of 100 independent noise sequences. They are reported for EMD-Shrinkage and for three Wavelets (Haar, Symmlet 4, Daubechies 4).

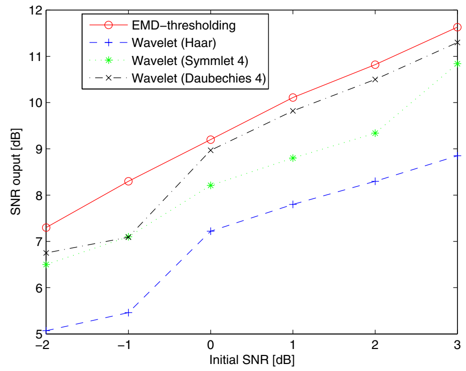

Fig. 6. Output SNR values obtained for different initial noise levels of signal ”b”. The results are the average of 100 independent noise sequences. They are reported for EMD-Shrinkage and for three Wavelets (Haar, Symmlet 4, Daubechies 4).