Advances in Adaptive Data Analysis
Vol. 2, No. 1 (2010) 1–16

World Scientific
www.worldscientific.com

VOICED SPEECH ENHANCEMENT BASED ON ADAPTIVE FILTERING OF SELECTED INTRINSIC MODE FUNCTIONS

KAIS KHALDI*,†,‡, MONIA TURKI-HADJ ALOUANE†,§
and ABDEL-OUAHAB BOUDRAA*,¶

*IRENav, Ecole Navale (EA 3634),
BCRM Brest, CC 600, 29240 Brest, France

† Unité Signaux et Systèmes,
Ecole Nationale d'Ingénieurs de Tunis,
BP 37, Le Belvédère 1002, Tunis, Tunisia
‡kais.khaldi@gmail.com
§m.turki@enit.rnu.tn
¶boudra@ecole-navale.fr

In this paper a new method for voiced speech enhancement combining the Empirical Mode Decomposition (EMD) and the Adaptive Center Weighted Average (ACWA) filter is introduced. Noisy signal is decomposed adaptively into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs). Since voiced speech structure is mostly distributed on both medium and low frequencies, the shorter scale IMFs of the noisy signal are beneath noise, however the longer scale ones are less noisy. Therefore, the main idea of the proposed approach is to only filter the shorter scale IMFs, and to keep the longer scale ones unchanged. In fact, the filtering of longer scale IMFs will introduce distortion rather than reducing noise. The denoising method is applied to several voiced speech signals with different noise levels and the results are compared with wavelet approach, ACWA filter and EMD–ACWA (filtering of all IMFs using ACWA filter). Relying on exhaustive simulations, we show the efficiency of the proposed method for reducing noise and its superiority over other denoising methods, i.e., to improve Signal-to-Noise Ratio (SNR), and to offer better listening quality based on a Perceptual Evaluation of Speech Quality (PESQ). The present study is limited to signals corrupted by additive white Gaussian noise.

Keywords: Voiced speech enhancement; Empirical Mode Decomposition; ACWA filter.

1. Introduction

The aim of noise reduction in speech is to lower the noise level without affecting the speech signal quality. In many speech communication applications, the

¶Corresponding author.

recorded and transmitted speech signals contain a considerable amount of acoustic background noise. Furthermore, with the growth of mobile communication applications, the problem of reducing the background noise has become increasingly important. Different strategies have been proposed for noise reduction, such as Wiener filter [Proakis and Manolakis (1996)] or subspace filtering [Hermus *et al.* (2007)]. These linear methods have attracted significant interests and investigations due to their easy design and implementation. However, these approaches are not very effective when signals contain sharp shapes or impulses of short duration. To overcome these limits, nonlinear approaches, such as wavelet analysis, have been proposed [Donoho (1995)]. However, the fixed basis functions limit the performance of the wavelets over particular class of nonstationary signals. Recently, a new data-driven method, called Empirical Mode Decomposition (EMD), has been introduced by Huang *et al.* [1998] for analyzing nonlinear and nonstationary signals. The EMD decomposes adaptively a signal into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs). The basis functions of EMD are derived from the signal itself and hence, the analysis is adaptive in contrast to traditional methods where the basis functions are fixed.

In this paper, a denoising scheme combining two adaptive methods and dedicated to voiced speech signals is proposed. The method is based on the EMD and the Adaptive Center Weighted Average (ACWA) filter [Lee (1980)] that both perform in time space. In our previous works [Khaldi *et al.* (2008a; 2008b)], the denoising method is based on the filtering of all IMFs extracted from the noisy signal. Since voiced speech signal energy is distributed over low and medium frequencies [Hermus *et al.* (2007)], the lower order of IMFs (high-frequency components) of the noisy signal is noise-contaminated [Weng *et al.* (2006); Boudraa and Cexus (2007)]. However, the longer scale IMFs (low- and medium-frequency components) corresponding to the most important structures of the signal are signal dominated. Therefore, filtering of these IMFs will introduce signal distortion rather than a noise reduction [Cexus (2006)]. The basic idea of the proposed method is to only filter the shorter scale IMFs (high-frequency components), which are noise dominated, and to keep the longer scale IMFs unchanged. This method is effective for voiced speech since the most important spectral features of voiced speech signal are distributed over medium and low frequencies [Hermus *et al.* (2007)]. Indeed, the power spectrum density of the voiced speech is very low for high frequencies. A criterion based on IMFs's energy is used to detect the shorter scale IMFs that contain much more noise than signal [Boudraa and Cexus (2007)]. These IMFs are filtered using ACWA filter [Lee (1980)], which operates adaptively in the time domain and does not require the stationarity and the whiteness of the signal. The proposed method is applied to voiced speech signals corrupted with additive white Gaussian noise. Comparisons with some denoising methods (ACWA filtering of all IMFs, wavelet denoising approach, and ACWA filtering of the noisy voiced signal) are performed.

2. EMD Basics

The EMD decomposes a given signal $x(t)$ into a set of IMFs through an iterative process called *sifting*; each one with a distinct time scale [Huang *et al.* (1998)]. 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 subbands 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 [Flandrin *et al.* (2004)]. 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 one extracted just before. The EMD does not use a pre-determined filter or basis functions, and it is a fully data-driven method [Huang *et al.* (1998)]. To be successfully decomposed into IMFs, the signal $x(t)$ must have at least two extrema, one minimum and one maximum. The IMFs are extracted using an algorithm called sifting process summarized as follows [Huang *et al.* (1998)]:

- identify all extrema of $x(t)$;
- interpolate between minima (resp. maxima), ending up with some envelope $e_{\min}(t)$ (resp. $e_{\max}(t)$);
- compute the average $m(t)$ $[(e_{\min}(t) + e_{\max}(t))/2]$;
- extract the detail $d(t)$ $[x(t) - m(t)]$; and
- iterate on the residual $m(t)$.

The signal $d(t)$ is considered a true IMF, if it satisfies the conditions (1) and (2). The result of the sifting is that $x(t)$ will be decomposed into a sum of C IMFs and a residual $r_C(t)$ such as the following:

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

where $\mathrm{IMF}_j(t)$ is the IMF of order j and, $r_C(t)$ is the residual.

3. Interest of ACWA Filter

Classically, the ACWA filter has been used in image enhancement applications [Lee (1980); Russo (1996)]. It can be also interesting and effective in the context of audio signal enhancement. As shown by Eq. (2), the ACWA filter operates in the time domain. In contrast to the classical filters, such as Wiener filter, all the parameters

are computed in time domain and, hence, transformation to frequency domain is not necessary. Besides, the noise variance is computed at all instants and the signal is enhanced sample by sample. The ACWA filtered signal $\tilde{x}(t)$ is described as follows [Lee (1980)]:

$$\tilde{x}(t) = \begin{cases} F_{\text{mean}} + K(\text{y}(t) - F_{\text{mean}}) & \text{if } F_{\text{var}} \geq \sigma^2 \\ F_{\text{mean}} & \text{otherwise,} \end{cases} \tag{2}$$

where

$$K = 1 - \frac{\sigma^2}{F_{\text{var}}}, \tag{3}$$

where F_{mean} and F_{var} denote, respectively, the average and the variance of the noisy signal $y(t)$ computed over a sliding window of size L, and σ^2 designates the variance of noise contained in the noisy signal $y(t)$. In order to show the effectiveness of this filter in the audio context, a comparative analysis between ACWA filter and Minimum Mean Square Error (MMSE) filter [Soon *et al.* (1998)] is presented.

In this work we consider the enhancement of speech sequences corrupted by additive white Gaussian noise. The noise level is fixed through the input Signal-to-Noise Ratio (SNR_{in}) to 2 dB. Figure 1 shows the superposition of the clean signal and the filtered signals obtained by the ACWA and the MMSE filters.

The comparative analysis of the three signals (Fig. 1) does not clearly show the superiority of the ACWA filter over the MMSE one. Therefore, we use the output SNR (SNR_{out}) and the Perceptual Evaluation of Speech Quality (PESQ)

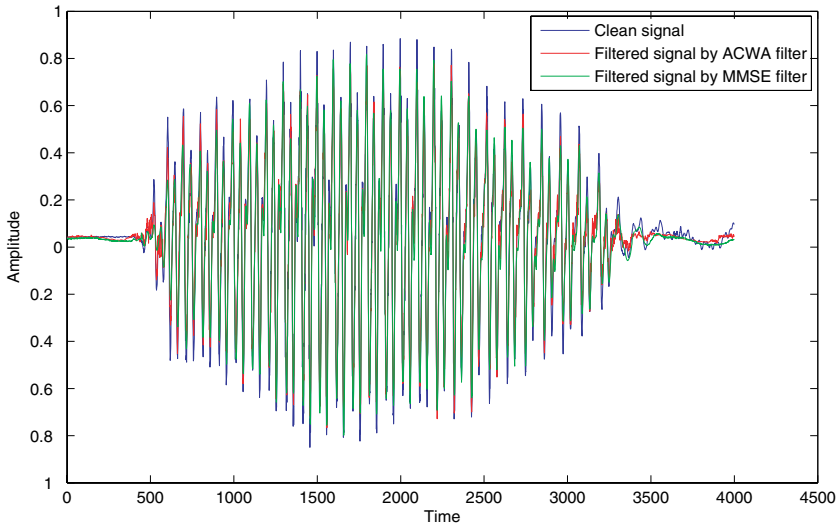


Fig. 1. Clean and filtered signals by the ACWA and the MMSE filters ($\text{SNR}_{\text{in}} = 2$ dB).

Table 1. Variations of the SNR_{out} and the PESQ over the SNR_{in}; relating to MMSE filter and the ACWA filter.

	MMSE filter		ACWA filter	
SNR input [dB]	SNR output [dB]	PESQ	SNR output [dB]	PESQ
−10	0.7	0.70	1.2	1.51
−8	1.53	0.91	2.01	1.72
−6	3.52	1.07	5.94	1.98
−4	5.00	1.2	7.98	2.07
−2	7.37	1.51	10.18	2.15
0	9.82	2.05	11.19	2.21
2	12.63	2.13	12.08	2.35
4	13.76	2.25	13.95	2.41
6	15.88	2.49	15.67	2.65
8	16.53	2.64	16.52	2.75
10	17.23	2.73	17.18	2.82

[Rix *et al.* (2001); ITU-T P.835 (2003)] criteria to quantify the speech enhancement quality obtained by the two filters. Table 1 presents the obtained results for different levels of the additive noise fixed through the SNR_{in}. These results show that for very low SNR_{in} values, the ACWA filter gives higher SNR_{out} than the MMSE filter. However, for all considered SNR_{in} values, the PESQ values given by the MMSE filter are higher than those related to the MMSE filter. The PESQ results confirm that the ACWA filter guarantees better listening quality of the enhanced speech than the MMSE filter.

4. Proposed Voiced Speech Denoising Approach

The proposed denoising method is illustrated by the scheme shown in Fig. 2. The noisy signal $y(t)$ is given by:

$$y(t) = x(t) + b(t), \tag{4}$$

where $x(t)$ corresponds to the clean voiced speech signal and $b(t)$ denotes an additive white Gaussian noise. The noisy signal is decomposed into a sum of IMFs by the EMD, such as:

$$y(t) = \sum_{j=1}^{C} \mathrm{IMF}_j(t) + r_C(t). \tag{5}$$

The denoising method consists in filtering by the ACWA filter a set of IMFs selected using an energy criterion [Boudraa and Cexus (2007)].

4.1. *IMFs selection*

The EMD filtering method relies on the basic idea that most important structures of the signal, such as voiced speech signal, are concentrated on medium and low

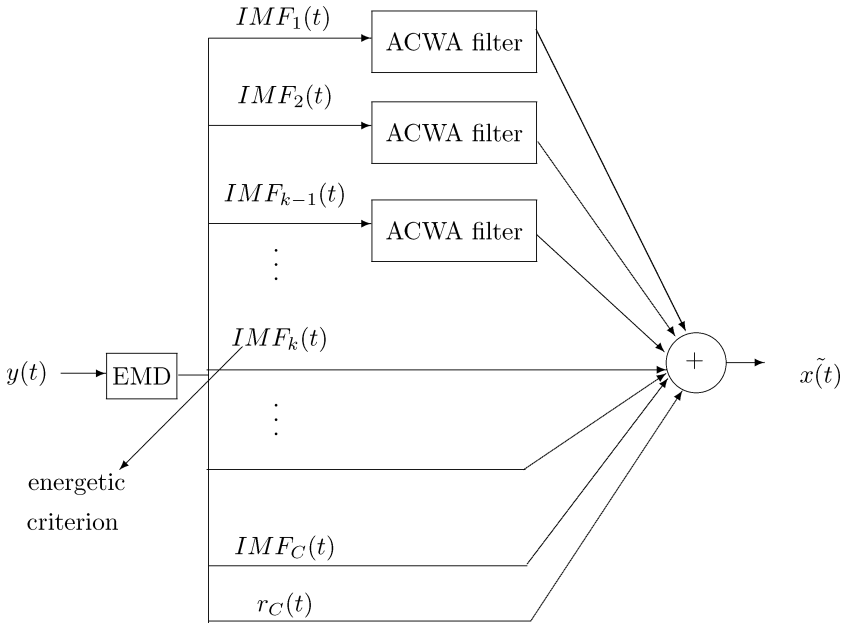


Fig. 2. Scheme of the proposed denoising approach.

frequencies, that correspond to longer scale IMFs [Weng *et al.* (2006); Flandrin *et al.* (2005)]. Therefore, the shorter scale IMFs of the noisy version are noise dominated, while the longer scale ones are signal dominated and their filtering can induce distortion of the reconstructed signal. According to this idea, there will be a mode, indexed by j_s, from which the energy distribution of the important structures of the signal overcomes that of the noise [Boudraa and Cexus (2007)]. Thus, a criterion based on energy density can be used [Wu and Huang (2004); Flandrin *et al.* (2005)].

From the observed signal $y(t)$, the objective is to find an approximation $\tilde{x}(t)$ to the original signal $x(t)$ that minimizes the Mean Square Error (MSE):

$$\mathrm{MSE}(x,\tilde{x}) \triangleq \frac{1}{N}\sum_{i=1}^{N}(x(t_i)-\tilde{x}(t_i))^2, \tag{6}$$

where $x = [x(t_1), x(t_2), \ldots, x(t_N)]^T$ and $\tilde{x} = [\tilde{x}(t_1), \tilde{x}(t_2), \ldots, \tilde{x}(t_N)]^T$. N is the length of the signal. Other distortion measures such as the Mean Absolute Error (MAE) can be used. Then, $y(t)$ is first decomposed using the EMD into $\mathrm{IMF}_j(t), j = 1, \ldots C$, and a residual $r_C(t)$, and finally $\tilde{x}(t)$ is reconstructed using $(C-k+1)$ selected IMFs starting from k to C (Eq. (7)).

$$\tilde{x}_k(t) = \sum_{j=k}^{C}\mathrm{IMF}_j(t) + \mathrm{r}_C(t), k = 2, \ldots, C. \tag{7}$$

The aim of the EMD filtering, which is carried out on time domain, is to find the index $k = j_s$ that minimizes the MSE$(x, \tilde{x})$. Note that Eq. (7) corresponds to a low-pass time-space filtering [Huang *et al.* (2006)]. In practice the MSE or the MAE cannot be calculated because the original signal $x(t)$ is unknown. In this paper we use a distortion measure called Consecutive MSE (CMSE) that does not require the knowledge of $x(t)$ [Boudraa and Cexus (2007)]. This quantity measures the squared Euclidean distance between two consecutive reconstructions of the signal. The CMSE is defined as follows [Boudraa and Cexus (2007)]:

$$\mathrm{CMSE}(\tilde{x}_k, \tilde{x}_{k+1}) \triangleq \frac{1}{N}\sum_{i=1}^{N}(\tilde{x}_k(t_i) - \tilde{x}_{k+1}(t_i))^2, \quad k = 1, \ldots, C-1, \tag{8}$$

$$\triangleq \frac{1}{N}\sum_{i=1}^{N}(\mathrm{IMF}_k(t_i))^2. \tag{9}$$

Thus, according to Eq. (9) the CMSE is reduced to the energy of the k^{th} IMF. It is also the classical empirical variance estimate of the IMF. Remark if $k = 1$, $\tilde{x}_k(t) = y(t)$. Finally, the index j_s is given by:

$$j_s = \underset{1 \le k \le C-1}{\mathrm{argmax}}\,[\mathrm{CMSE}(\tilde{x}_k, \tilde{x}_{k+1})], \tag{10}$$

where $\tilde{x}_k$ and $\tilde{x}_{k+1}$ are signals reconstructed starting from the IMFs indexed by k and $(k+1)$, respectively. The CMSE criterion allows to identify the IMF order where there is the first significant change in energy. This empirical fact is derived from extensive experiments and simulations [Boudraa and Cexus (2007)]. Once the index j_s is calculated, the IMFs of order $j < j_s$ are filtered and those of order $j \ge j_s$ are not processed.

4.2. *ACWA filtering*

The shorter scale $(j_s - 1)$ IMFs, which are hidden beneath noise, are filtered by the ACWA filter performing in the time space. In fact, each IMF_j(t) of order $(j < j_s)$ is assumed to be a noisy version of the data $f_j(t)$. So it can be expressed as:

$$\mathrm{IMF}_j(t) = f_j(t) + b_j(t). \tag{11}$$

An estimate $\tilde{f}_j(t)$ of $f_j(t)$ is given by:

$$\tilde{f}_j(t) = \Gamma[\mathrm{IMF}_j(t)], \tag{12}$$

where $\Gamma[\mathrm{IMF}_j(t)]$ is a temporal processing [Boudraa and Cexus (2007)] corresponding in this case to ACWA filter. The noise level $\tilde{\sigma_j}$ of IMF_j can be computed as follows [Teukolsky *et al.* (1992); Boudraa and Cexus (2006)]:

$$\tilde{\sigma_j} = 1.4826 \times \mathrm{Median}\{|\mathrm{IMF}_j(t) - \mathrm{Median}\{\mathrm{IMF}_j(t)\}|\}. \tag{13}$$

The proposed denoising approach is described in four steps as follows:

Input: Noisy voiced speech: $y(t)$.
Output: Denoised voiced speech: $\tilde{x}(t)$.
Initialization:
$h(t) \leftarrow y(t)$
Step A: Decompose $y(t)$, by EMD, into j IMFs, $j \in \{1, \ldots, C\}$, and the residual $r_C(t)$.
Step B: Calculate the energy of each IMFs, and find the index j_s using Eq. (10).
Step C: Denoising the shorter scale $(j_s - 1)$ IMFs using relations (3) and (4).
Step D: The denoised signal, $\tilde{x}(t)$, is reconstructed as follows:

$$\tilde{x}(t) = \sum_{j=1}^{j_s-1} \tilde{f}_j(t) + \sum_{j=j_s}^{C} \mathrm{IMF}_j(t) + r_C(t). \tag{14}$$

5. Results

The proposed noise reduction method is tested on voiced speech signals corrupted by varying additive white Gaussian noise levels, fixed through the $\mathrm{SNR_{in}}$. Four clean voiced speech signals vowels /o/, /a/, /e/ and /i/ (Fig. 3) pronounced by a male speaker are analyzed.

These signals are corrupted by an additive white Gaussian noise with SNR values ranging from -10 dB to 10 dB. The results of the proposed scheme are compared

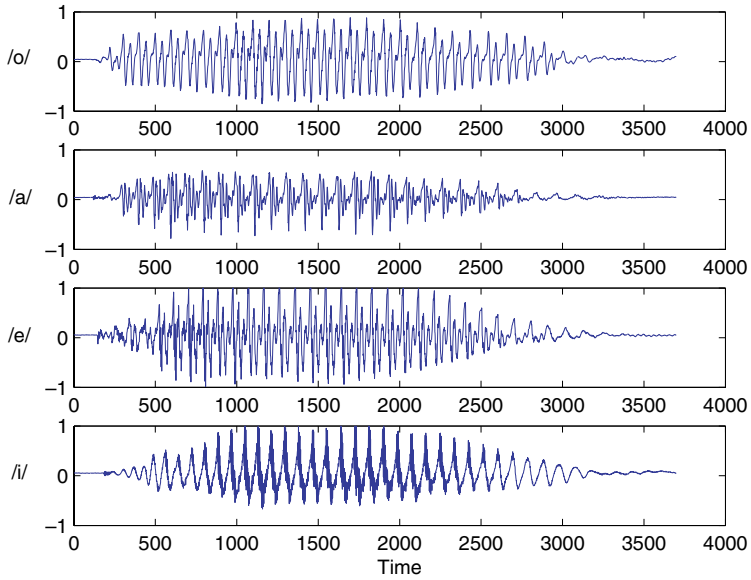


Fig. 3. Original signals /o/, /a/, /e/ and /i/.

with those of three methods: ACWA filtering of all IMFs (EMD–ACWA), denoising based on wavelet decomposition [Khaldi *et al.* (2008a;2008b)], and ACWA filtering of the noisy voiced signal. The performance evaluation is based on the $\mathrm{SNR}_{\mathrm{out}}$ and the PESQ measures. For each $\mathrm{SNR}_{\mathrm{in}}$ value, 100 independent noise realizations are generated and averaged values of the $\mathrm{SNR}_{\mathrm{out}}$ and the PESQ are computed. Noisy versions of the original signals corresponding to $\mathrm{SNR}_{\mathrm{in}} = 2\,\mathrm{dB}$ are shown in Fig. 4.

For illustration, Fig. 5 shows that the EMD decomposes the noisy signal /o/ into ten IMFs and a residual. According to this decomposition, we can see that from the fourth IMF, the original signal components are more dominant than the noise components. This finding is well verified based on CMSE criterion.

Indeed, Fig. 6 shows that for the sequence /o/, the maximum of CMSE points out at the fourth IMF. Figure 6 shows the plots of the CMSE values *versus* the number of IMFs for the four signals. Each curve is characterized by only one maximum corresponding to the index j_s. Table 2 summarizes for each signal, the number of IMFs given by the EMD decomposition; and the index j_s corresponding to the largest CMSE or IMF energy. The second stage of the proposed method consists in filtering the $(j_s - 1)$ shorter scale IMFs using the ACWA filter. The size, L, of the sliding window of ACWA filter is set to 511. Such setting is justified by the results shown in Fig. 7 where are displayed the variations of the $\mathrm{SNR}_{\mathrm{out}}$ *versus* the L values.

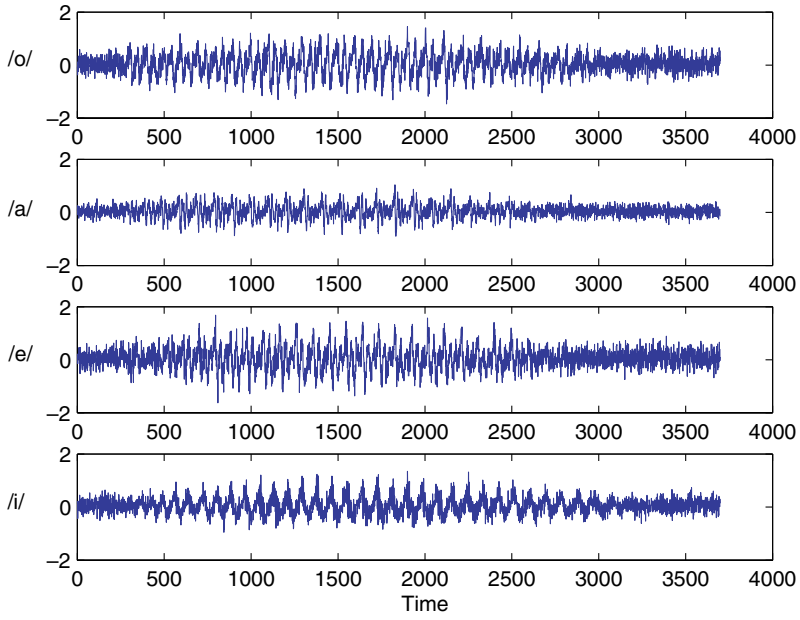


Fig. 4. Noisy versions of signals /o/, /a/, /e/ and /i/ ($\mathrm{SNR}_{\mathrm{in}} = 2\,\mathrm{dB}$).

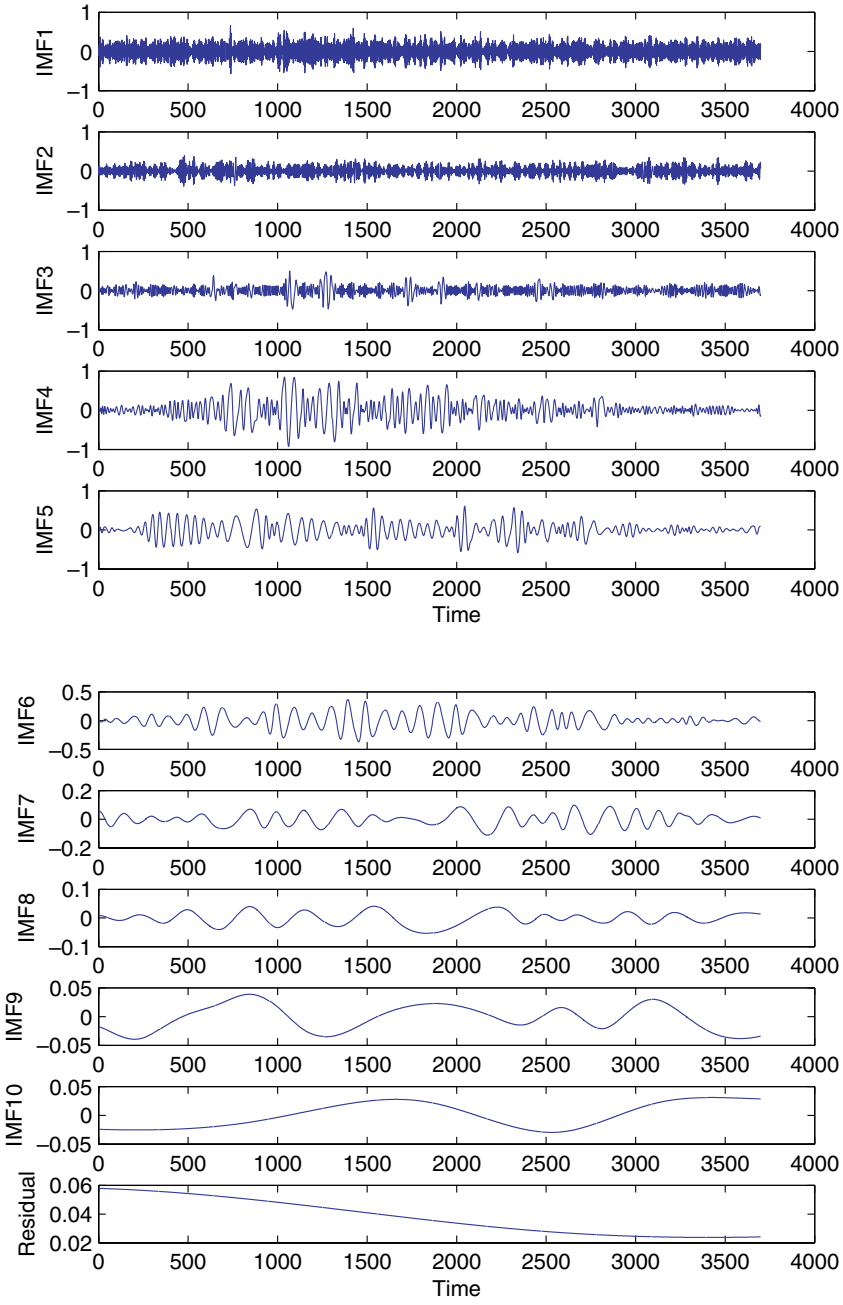


Fig. 5. Decomposition of noisy signal /o/ by EMD ($SNR_{in} = 2\,dB$).

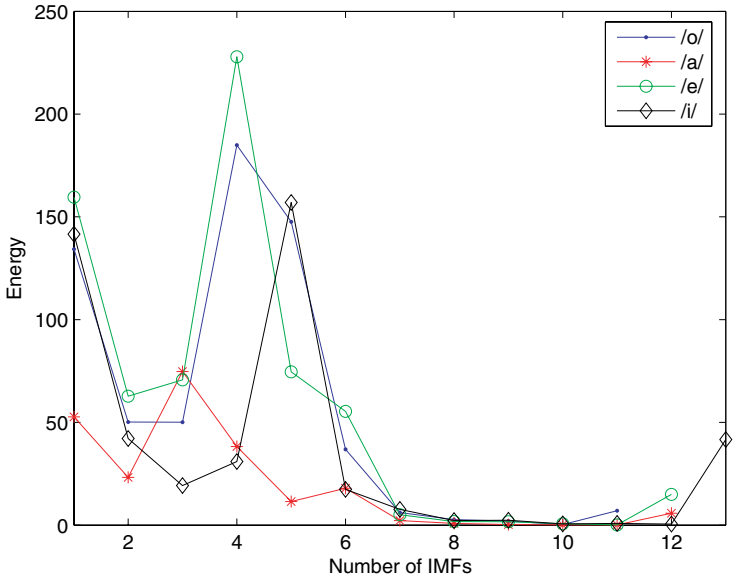


Fig. 6. Variations of CMSE (energy) values *versus* the number of IMFs for the four noisy signals.

Table 2. C and j_s values of each signal.

Signals	/o/	/a/	/e/	/i/
C	10	11	11	12
j_s	4	3	4	5

Figure 7 shows that for the three considered values of $\mathrm{SNR_{in}}$, the $\mathrm{SNR_{out}}$ remains almost constant for $L \geq 511$.

Denoising results obtained by the proposed method, the ACWA filtering of the noisy signal, the ACWA filtering of the all IMFs of the noisy signal (EMD–ACWA), and a denoising based on the wavelet (db8) thresholding [Khaldi *et al.* (2008a;2008b)], are shown in Fig. 8 for an $\mathrm{SNR_{in}} = 2\,\mathrm{dB}$. In fact, we choose db8 with a hard threshold as a tool of comparison, because it gives good results compared to the other wavelets. A careful comparative examination of the signals, as shown in Figs. 3 and 8, shows that the proposed method performs better than the other three methods in terms of noise reduction.

This conclusion is confirmed by the $\mathrm{SNR_{out}}$ values listed in Table 3. For all voiced speech signals the SNR gain achieved by the proposed method is the highest.

These findings are confirmed by the results shown in Fig. 9. It is shown that for the four signals the proposed method performs remarkably better than the EMD–ACWA and the other methods. The SNR improvement achieved by the proposed method varies from 3.4 dB to 17.9 dB. For very lower $\mathrm{SNR_{in}}$ values, we still observe

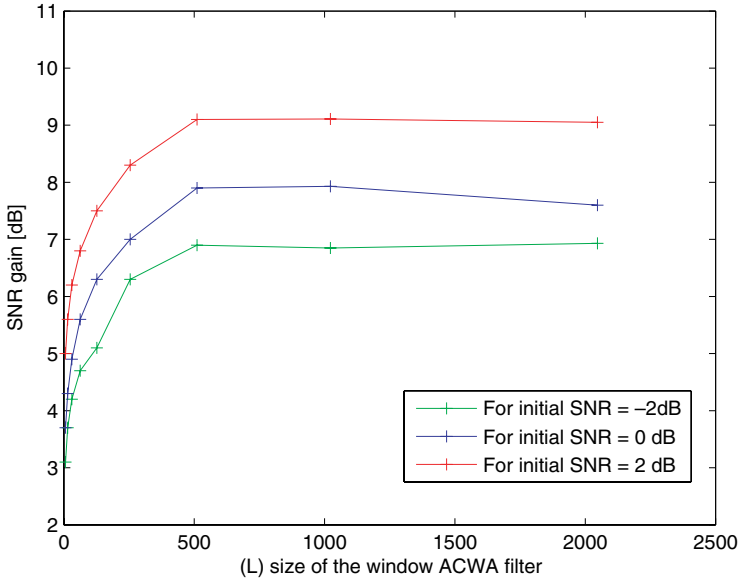
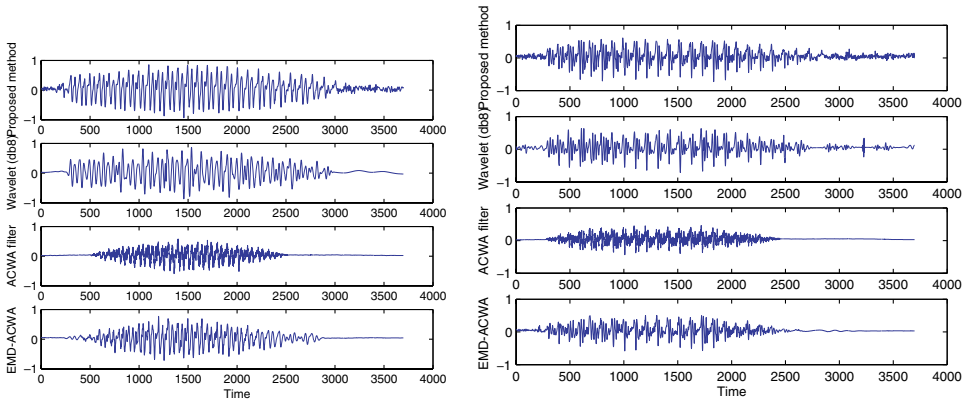


Fig. 7. Variations of SNR_{out} values *versus* L ($SNR_{in} = -2$ dB, 0 dB and 2 dB).

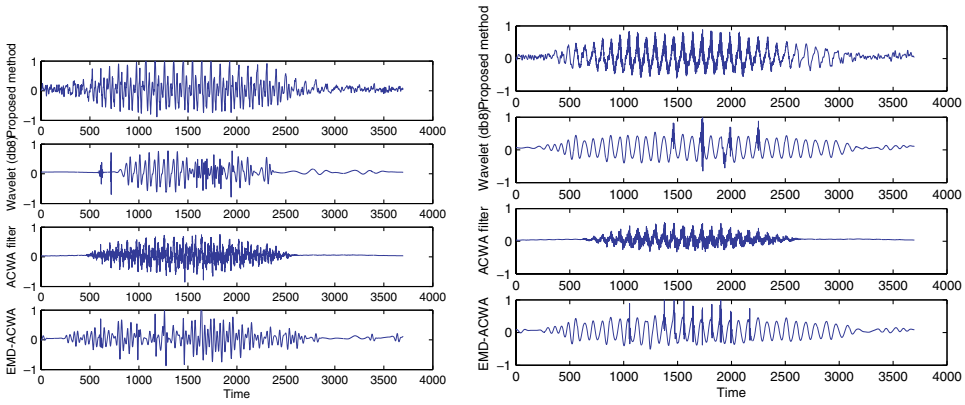
the effectiveness of the proposed method in removing the noise components. Indeed, the SNR improvement is all the more high since the SNR_{in} is low.

When listening to the enhanced speech signals, the proposed method produces lower residual noise and noticeably less speech distortion for all the signals. This result is confirmed by the PESQ results shown in Fig. 10. These results



(a) Denoising of noisy signal /o/. (b) Denoising of noisy signal /a/.

Fig. 8. Enhanced signals obtained by the proposed method, Wavelet (db8), ACWA filter, and EMD–ACWA ($SNR_{in} = 2$ dB).



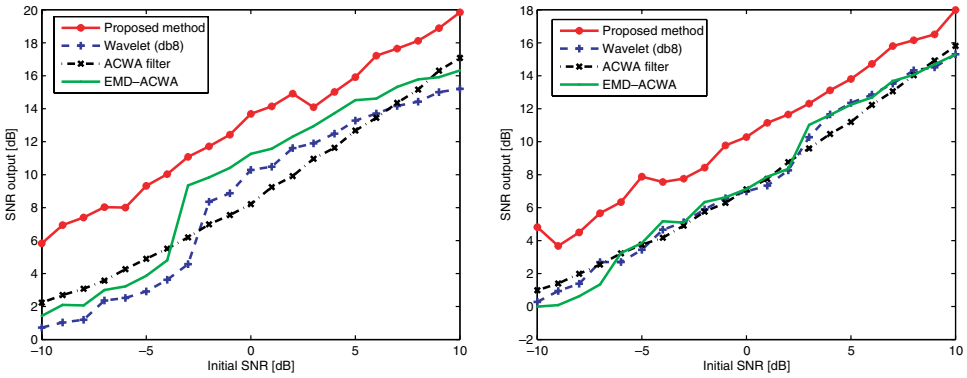
(c) Denoising of noisy signal /e/. (d) Denoising of noisy signal /i/.

Fig. 8. (*Continued*)

Table 3. Denoising results, based on the SNR_{out}, of four noisy voiced different signals ($SNR_{in} = 2$ db).

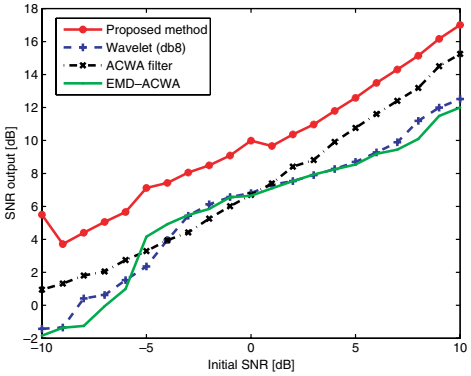
Noisy signals (SNR = 2 dB)	/o/	/a/	/e/	/i/
Proposed method	14.82	11.87	10.55	9.44
EMD–ACWA	11.94	7.87	7.41	5.23
Wavelet (db8)	11.38	7.85	7.40	5.24
ACWA filter	9.80	8.04	7.91	7.31

demonstrate that our approach gives a significant enhancement in listening quality as the improvement of the PESQ values is high. Indeed, the obtained results also show that it is more efficient to apply the ACWA filter to selected IMFs of the noisy signal than to the all IMFS.

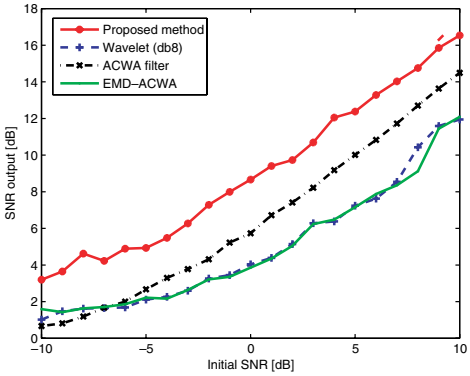


(a) Gain in SNR for noisy version of /o/. (b) Gain in SNR for noisy version of /a/.

Fig. 9. Variations of SNR_{out} *versus* SNR_{in} for signals /o/, /a/, /e/, and /i/. The results are the average of 100 noise realizations. The reported results correspond to proposed method, Wavelet (db8), ACWA filter, and the EMD–ACWA.

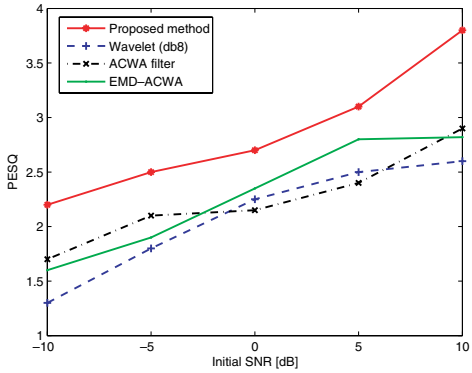


(c) Gain in SNR for noisy version of /e/.

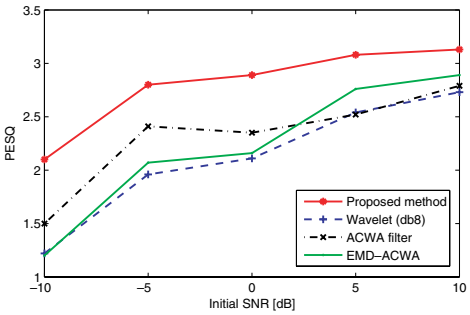


(d) Gain in SNR for noisy version of /i/.

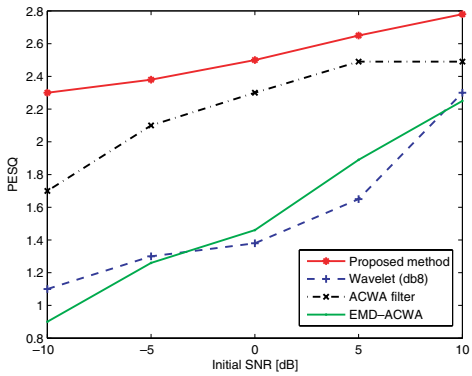
Fig. 9. (*Continued*)



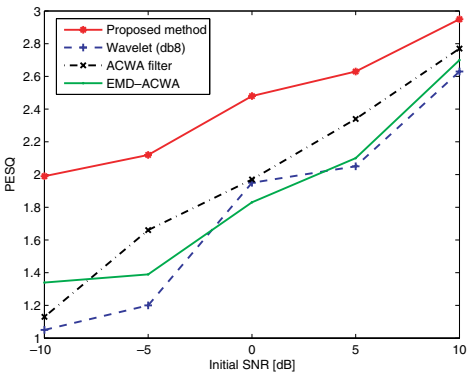
(a) PESQ for noisy version of /o/.



(b) PESQ for noisy version of /a/.



(c) PESQ for noisy version of /e/.



(d) PESQ for noisy version of /i/.

Fig. 10. Variations of PESQ values *versus* $\mathrm{SNR_{in}}$ for the signals /o/, /a/, /e/, and /i/. The results are the average of 100 noise realizations. The reported results correspond to proposed method, Wavelet (db8), ACWA filter, and the EMD–ACWA.

6. Conclusion

In this paper, a new voiced speech enhancement method is presented. To lower the noise level, two effective and powerful methods, pre-filtering by EMD and ACWA filtering, are combined. Obtained results for denoising voiced 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 wavelet approach, the ACWA filter, and the EMD–ACWA method. In addition, the PESQ criterion confirms that the proposed method offers a much better listening quality than the other methods.

Acknowledgments

This work is partially supported by grants from CMCU (Comité Mixte de Coopération Universitaire Franco-Tunisienne).

Appendices

In this appendix, we give brief descriptions of the quality measures used. **Input Signal-to-Noise Ratio ($\mathbf{SNR_{in}}$):** The input Signal to Noise Ratio ($\mathbf{SNR_{in}}$) is given by:

$$\mathrm{SNR}_{\mathrm{in}} = 10\log_{10}\frac{\sum_{t=1}^{T}(x(t))^2}{\sum_{t=1}^{T}(y(t)-x(t))^2}, \tag{A.1}$$

where x and y are, respectively, the clean and the noisy signals.

Output Signal-to-Noise Ratio ($\mathbf{SNR_{out}}$): The $\mathbf{SNR_{out}}$ is very sensitive to the time alignment of the original and distorted signals. The $\mathbf{SNR_{out}}$ is measured as:

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}\frac{\sum_{t=1}^{T}(\tilde{x}(t))^2}{\sum_{t=1}^{T}(x(t)-\tilde{x}(t))^2}, \tag{A.2}$$

where $\tilde{x}$ is the reconstructed signal.

Perceptual Evaluation of Speech Quality (PESQ): The PESQ measure is the most complex to compute, and it is recommended by ITU-T for speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codec [ITU-T P.835 (2003)]. The note refers PESQ values type MOS, in the form of a scalar between −0.5 and 4.5.

References

Boudraa, A. O. and Cexus, J. C. (2006). Denoising via empirical mode decomposition. *Proc. IEEE ISCCSP*, 1–4.

Boudraa, A. O. Cexus, J. C. and Saidi, Z. (2004). EMD-based signal noise reduction. *Int. J. Signal Process.* **1**, 1: 33–37.

Boudraa, A. O. and Cexus, J. C. (2007). EMD-based signal filtering. *IEEE Trans. Instrum. Meas.*, **56**, 6: 2196–2202.

Cexus, J. C. (2006). Analyse des signaux non-stationnaires par Transformation de Huang, Operation de Teager–Kaiser, et Transformation de Huang–Teager (THT). *PhD Thesis, University of Rennes I.*

Deger, E., Islam Molla, K., Hirose, K., Minemastu, N. and Hasan, K. (2007). Speech enhancement using soft thresholding with DCT-EMD based hybrid algorithm. *Proc. EUSIPCO*, 1–5.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, **41**: 613–627.

Flandrin, P. Goncalves, P. and Rilling, R. (2005). EMD equivalent filter banks, from interpretation to applications. *Hilbert–Huang Transform and Its Applications*, 57–73, ed. N. E. Huang and S. S. P. Shen, World Scientific.

Flandrin, P., Rilling, G. and Goncalves, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Proc. Lett.*, **11**, 2: 112–114.

Hermus, K., Wambacq, P. and Van Hamme, H. (2007). A review of signal subspace speech enhancement and its application to noise robust speech recognition. *EURASIP J. Adv. Signal Process.*, **2007**: 1–15.

Huang, N. E., Brenner, M. J. and Salvino, L. (2006). Hilbert–Huang transform stability spectral analysis applied to flutter flight test data. *AIAA J.*, **44**, 4: 772–786.

Huang *et al.* (1998). The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc.*, **454**: 903–995.

ITU-T P.835. (2003). Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, *ITU-T Recommendation.*

Khaldi, K., Boudraa, A. O., Bouchikhi, A., Turki-Hadj Alouane, M. and Diop, E. H. S. (2008a). Speech signal noise reduction by EMD. *Proc. IEEE ISCCSP*, 1–4.

Khaldi, K., Boudraa, A. O., Bouchikhi, A. and Turki-Hadj Alouane, M. (2008b). Speech Enhancement via EMD. *EURASIP J. Adv. Signal Process.*, **2008**: 1–8.

Lee, J. S. (1980). Digital image enhancement and noise filtering by using local statistics. *Pattern Anal. Mach. Intell.*, **2**, 4: 165–168.

Proakis, J. G. and Manolakis, D. G. (1996). *Digital Signal Processing*: *Principles, Algorithms, and Applications*, 3rd edn. Prentice-Hall.

Rix, A., Beerends, J., Hollier, M. and Hekstra, A. (2001). Perceptual evaluation of speech quality (PESQ) — A new method for speech quality assessment of telephone networks and codecs. *Proc. IEEE ICASSP*, **4**: 749–752.

Russo, F. (1996). Nonlinear fuzzy filters: An overview. *Proc. EUSIPCO*, 257–260.

Soon, I. Y., Koh, S. N. and Yeo, C. K. (1998). Noisy speech enhancement using discrete cosine transform. *Speech Commun.*, **24**, 3: 249–257.

Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C*: *The Art of Scientific Computing*, 2nd edn. W.H. Press.

Weng, B., Blanco-Velasco, M. and Barner, K. E. (2006). ECG Denoising based on the Empirical Mode Decomposition. *Proc. IEEE EMBS*, 1–4.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. Roy. Soc. Lond. Ser. A*, **460**: 1579–1611.

Wu, Z. and Huang, N. E. (2005). Statistical significance test of intrinsic mode functions in *Hilbert–Huang Transform and Its Applications*, 103–147, ed. N. E. Huang and S. S. P. Shen, World Scientific.