



Nonlocal means estimation of intrinsic mode functions for speech enhancement

Sagar Reddy VUMANTHALA^{1,*} , Bikshalu KALAGADDA² 

¹Department of Electronics and Communication Engineering,

VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

²Department of Electronics and Communication Engineering, Kakatiya University, Warangal, India

Received: 12.01.2019

Accepted/Published Online: 21.08.2019

Final Version: 27.01.2020

Abstract: The main aim of this paper is to introduce a new approach to enhance speech signals by exploring the advantages of nonlocal means (NLM) estimation and empirical mode decomposition. NLM, a patch-based denoising method, is extensively used for two-dimensional signals like images. However, its use for one-dimensional signals has been attracting more attention recently. The NLM-based approach is quite useful for removing low-frequency noises based on nonlocal similarities present among samples of the signal. However, there is an issue of under averaging in the high-frequency regions. The temporal and spectral characteristics of the speech signal are changing markedly over time. Thus NLM is conventionally not effective to remove the noise components from the speech signal, unlike image denoising. To address this issue, initially, the speech signal is decomposed into oscillatory components called intrinsic mode functions (IMFs) by using a temporal decomposition technique known as the sifting process. Each IMF represents signal information at a certain scale or frequency band. The IMFs do not have abrupt power spectral changes over time. The decomposed IMFs are processed using NLM estimation based on nonlocal similarities for better speech enhancement. The simulation result shows that the proposed method gives better performance in terms of subjective and objective quality measures. Its performance is evaluated for white, factory, and babble noises at different signal to noise ratios.

Key words: Speech enhancement, denoising, nonlocal means, empirical mode decomposition

1. Introduction

In the last few decades, removing noise from noisy speech signals has become important in the area of speech signal processing. The main aim of the speech enhancement process is to improve the quality of speech. Speech enhancement also underlies plenty of applications such as speech and speaker recognition, human-machine interactive systems, voice activity detection, and acoustic emotion identification. The challenging task in the speech enhancement process is to estimate the noise statistics, especially in nonstationary real-time environments. Generally, the speech signal is degraded by noises present at different levels like the recording environment and communication channel [1]. The elimination of noise components from a noisy speech signal is essential but also a challenging issue for improving the quality and intelligibility. To solve these problems, it is necessary to improve the quality of the speech signal, using suitable speech enhancement algorithms based on applications.

In the literature, over the years, researchers have implemented various speech enhancement algorithms. Speech enhancement algorithms are classified based on statistical noise estimation. The most commonly used speech enhancement technique was subtractive, by Weiss et al. in 1974. In this technique, the basic approach

*Correspondence: vsagarreddy1990@gmail.com

is to estimate the noise spectrum from nonvoice regions in a corrupted speech signal. Then the estimated noise spectrum is subtracted from the noisy speech to obtain the noise-free speech signal [2]. The efficiency of these approaches is based on the accuracy with which the nonvoice regions are detected and also through vigorous estimation of the noise spectrum [3, 4]. Wide variant improved versions of spectral subtraction have been developed by researchers such as Virag [5, 6]. Even though they have different variants, these techniques are limited due to the musical noise present in the enhanced speech signal.

The Wiener filter is another denoising technique to suppress the noise in a noisy speech [7, 8]. This approach is based on minimizing the mean square error (MSE) between the estimated and original signal magnitude spectra. These traditional methods are used because of their easy design and implementation, but they do not perform well in the case of speech, where the signal is nonlinear and nonstationary.

Time-frequency analysis techniques like the wavelets approach have also been adopted for speech enhancement [9, 10]. In wavelet transform approaches, the denoising method is mainly based on the correlation of wavelet coefficients, modulus maximum, and threshold denoising. Even though it has several advantages, this technique is limited by fixed basis functions and these do not necessarily match for all signals in real time.

Recently, a new data-driven technique to analyze nonstationary and nonlinear signals has been proposed by Hung et al. [11]. In this method, the basis functions are derived from the signal itself, unlike other traditional methods in which the basis functions are fixed. The EMD decomposes the given signal into a finite set of adaptive basis functions named intrinsic mode functions (IMFs). The estimated signal is reconstructed from the few IMFs that are signal dominated based on an energy criterion [12]. However, if the speech signal is corrupted with speech-like noise, these decomposition methods are also not effective for removing unwanted signal components.

The nonlocal means (NLM) estimation technique is extensively used for denoising signal-like images and electrocardiography (ECG) signals [13]. NLM is a very successful data adaptive image denoising method introduced by Buades et al. [14]. It is an efficient algorithm to remove the noise when signal samples contain nonlocal similarities among them. This algorithm is not effective for a nonstationary signal like speech because its power spectrum changes over time [15]. To overcome these limitations, first, we decompose the signal into some intrinsic mode functions by using empirical mode decomposition (EMD). Each IMF represents signal information at a certain scale or frequency band and there are no abrupt power spectral changes over time. Then each IMF is processed through NLM estimation based on nonlocal similarities for effective speech enhancement. Finally, the enhanced speech signal is reconstructed from the processed IMFs.

The rest of this paper is organized as follows. Section 2 describes the proposed method for speech enhancement by using NLM estimation of IMFs. The material to understand the EMD and NLM is also discussed in section 2. Section 3 describes the baseline techniques used for comparison of the proposed method. Section 4 shows the experimental results, and the method is compared with the three existing speech enhancement techniques for various noises with different SNRs. Finally, Section 5 concludes the paper.

2. The proposed method

Figure 1 shows a block diagram of the proposed method to enhance the speech signal. In this method, the noisy speech signal is processed through the following steps in order to get an enhanced speech signal.

In the proposed method, the speech enhancement is achieved by NLM estimation of each IMF. In the following subsections, the process flow of EMD, NLM estimation of IMFs, and selection of parameters is discussed briefly.

1. The noisy speech signal is decomposed into k number of intrinsic mode functions (IMFs) using EMD, from high frequency oscillatory components to low frequency components. Each IMF satisfies two basic conditions.
2. Each IMF is processed through NLM estimation based on similarities present among the samples to remove the noise components.
3. Finally, the NLM estimated signals obtained from each IMF are combined to obtain the enhanced signal.

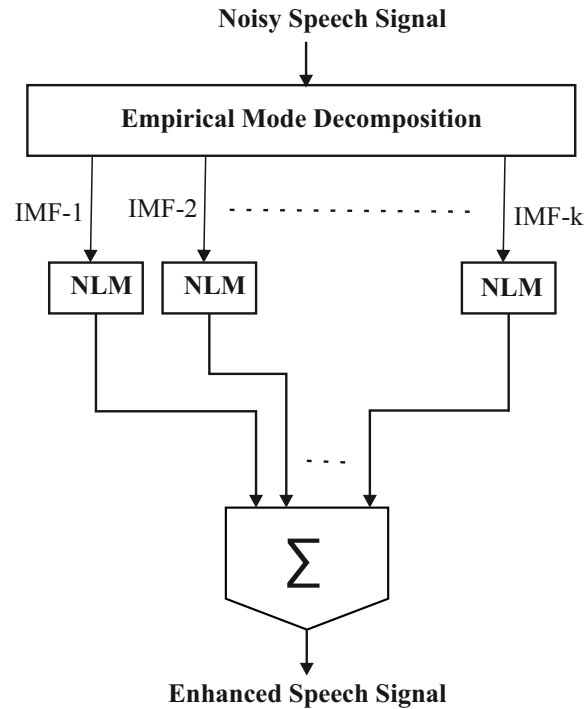


Figure 1. Block diagram representing the proposed method for speech enhancement.

2.1. Empirical mode decomposition

EMD is a data-driven method to analyze nonstationary and nonlinear data. The signal is analyzed between two consecutive extrema (minima and maxima). In this method, a signal $x(t)$ is decomposed into oscillatory components called intrinsic mode functions (IMFs). Each one with a distinct time scale is processed using a temporal decomposition method called a sifting process. These IMFs are not predefined like Fourier and wavelet transforms but are adaptively extracted from the input signal itself. The essence of this method is to identify the intrinsic oscillatory modes by their characteristics time scales in data empirically. Each intrinsic mode function satisfies two basic conditions: (i) the number of zero crossings and the number of extrema must be the same or differ at most by one. (ii) At any point, the mean of the envelope defined by the local maxima and minima is always zero. The IMFs are oscillatory functions and do not have a DC component.

Figure 2 shows the process flow of the EMD algorithm. According to the definition of IMF, the decomposition method can simply use envelopes defined by local extremes of $x(n)$. The upper and lower envelopes should cover all the data between them. Their local mean is considered $m(n)$ and the extract detailed signal $h(n)$ by subtracting $m(n)$ from $x(n)$. If $h(n)$ does not satisfy the stoppage criteria, then the process is recursive

and $h(n)$ is input for the next step. Otherwise, $h(n)$ is considered an n^{th} IMF. The residue or remaining signal is $r(n) = x(n) - h(n)$. The signal can be reconstructed by combining n IMFs and residual $r(n)$.

$$x(n) = \sum_{i=1}^{\infty} IMF_i[n] + r(n)$$

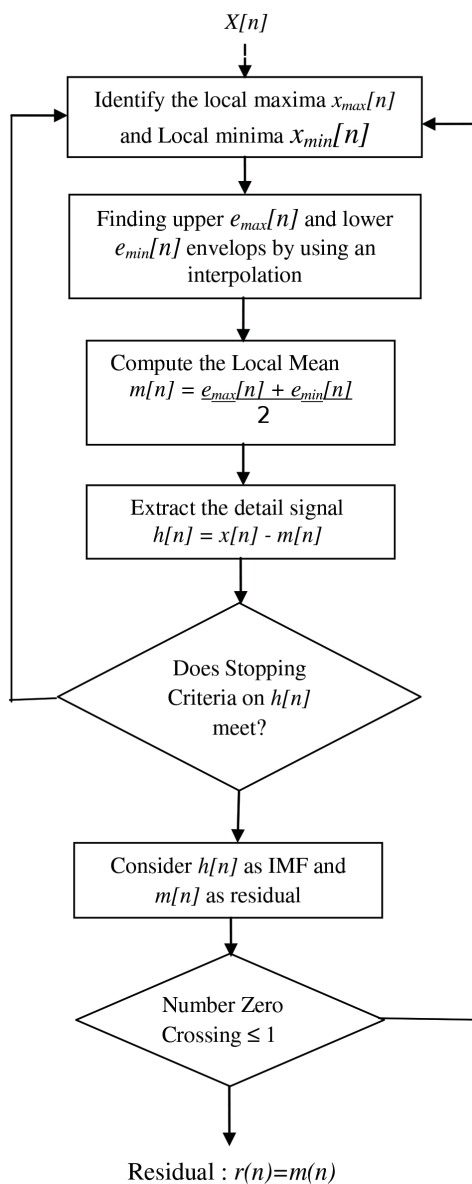


Figure 2. Process flow of EMD algorithm.

The above sifting process has two effects: (i) eliminating riding signals and (ii) smoothing uneven magnitudes. The first condition is necessary to get a meaningful instantaneous frequency. If the neighboring wave amplitudes are too large a discrepancy, then the second condition is also necessary. When carried out to

the extreme, it could decimate the physically meaningful amplitude fluctuations. Therefore, sifting should be applied with care in case of carrying the process to an extreme, and so we need to find the stopping criterion to guarantee that the IMFs retain enough physical sense of both amplitude and frequency modulation. This can be achieved by limiting the size of standard deviation (SD), computed from two consecutive sifting processes, which results in

$$SD = \sum_{t=0}^T \left[\frac{|(h_{1(k-1)}(n) - h_{1(k)}(n))|^2}{h_{1(k-1)}^2(n)} \right]$$

Usually the typical value of SD is set between 0.2 and 0.3.

2.2. Nonlocal means estimation

The NLM technique estimates the actual signal from the corrupted signal by exploiting the nonlocal similarities present among the samples of a signal. The NLM method is a very successful data adaptive image denoising method introduced by Buades. This technique has been proved to be very efficient for image denoising. Figure 3 shows an illustration of NLM parameters. To the best of our knowledge, the NLM technique has not been used extensively for speech enhancement yet. NLM is a patch-based denoising technique; denoising of a given patch is obtained as a weighted average of neighborhood patches, with the weights proportional to the patch similarity. In the NLM denoising technique, for each sample, $u(i)$ is the estimated sample and $\hat{u}(i)$ is a weighted sum of values at sample point m . A sample speech signal $U = \{u(i), i \in Z\}$.

For each sample $u(i)$ consider a set called searched region S_i with the size of $2K+1$.

$$S_i = \{u(i - K), u(i - K - 1), \dots, u(i - 1), u(i), u(i + 1), \dots, u(i + K - 1), u(i + K)\}$$

Here $u(i)$ is considered a central sample.

The sample and the set of neighborhood around it are defined by a window M_i of size $M= 2L+1$.

$$M_i = \{u(i - L), u(i - L - 1), \dots, u(i - 1), u(i), u(i + 1), \dots, u(i + L - 1), u(i + L)\}$$

NLM estimates $\hat{u}(i)$ as the weighted sum of samples in noisy speech within the prescribed search region S_i as follows:

$$\hat{u}(i) = \frac{1}{W_i} \sum_{m \in S_i} W(i, m) x(m)$$

A weight for each sample in the search region S_i is computed by finding the nonsimilarity present in the neighborhood with respect to the sample points $x(i)$ and $x(m)$.

$$W(i, m) = \exp \left(- \frac{\sum_{m=-L}^L [x(i) - x(i + m)]^2}{2N\lambda^2} \right)$$

Here λ represents the bandwidth parameter that controls the amount of smoothing applied. The patch width (N) selects the scale on which patches are compared and normalized in order to get the required weight value.

W_i represents the normalized weight as the summation of all weight values at sample point i and is computed as follows:

$$W_i = \sum_{m \in S_i} W(i, m)$$

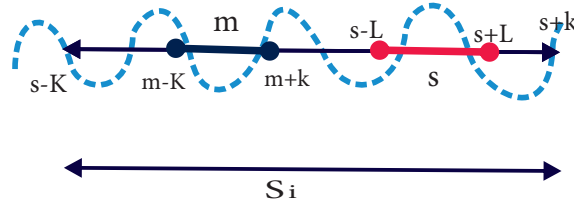


Figure 3. Illustration of NLM parameters. Two patches are compared in S_i .

2.3. Parameter selection

The efficiency of the proposed method depends on the number of decomposition levels and NLM tuning parameters. The performance of NLM estimation depends on key parameter selection like bandwidth (λ), neighborhood window size (M), and patch size (N).

The bandwidth λ : It is the primary parameter that controls the amount of smoothing applied. A small value of λ results in insufficient averaging due to too much impudence in different deweighting patches. For a large value of λ , it results in distortion, which makes dissimilar patches appear similar. The optimal value of λ is data dependent and noise-level dependent. To obtain an optimal value of λ , Ville and Kocher [16] have used SURE criteria, where an appropriate choice of lambda is 0.3σ to 0.8σ . σ is represented as the noise standard deviation of the corresponding IMF. In this method, the value of λ is selected as 0.4σ .

Neighborhood window size (M): The performance efficiency of a window is directly proportional to the size of the neighborhood window. For a large value of window size, it results in better averaging, but it is difficult to find similar windows. Usually, the properties of speech do not change much during the segment of 10–20 ms. Therefore, a neighborhood window size (M) of 80 to 160 is appropriate for speech signals as the sampling frequency of 8 kHz is taken. In this method, the value of M is selected as 80 empirically since better performance has been observed.

Patch size (N): The patch size selects the scale on which the patches are to be compared, and it should generally be of the size of the features of interest. With an increase in the size of the patch window, it is easier to find a similar window, but it could affect the chance of a false match and also increase the run time of the algorithm. Through observation, the optimal value can be suggested between 8 and 12 (1–2 ms) for speech signals. In this method, the value of N is selected as 8 empirically.

Number of IMFs: The noisy speech signal is decomposed into 2 IMFs after 15 iterations using EMD. The improvements in the average values of SegSNR, PESQ, BAK and OVL objective measures with different numbers of IMFs are shown in Figures 4a, 4b, 4c, and 4d, respectively. The best results are achieved with two IMFs.

2.4. Final speech enhancement by NLM estimation of IMFs

The temporal and spectral characteristics of the speech signal change markedly over time. NLM is a very successful data adaptive method to remove the noise when signal samples contain nonlocal similarities among them. This algorithm is not effective for a nonstationary signal like speech because its power spectrum changes

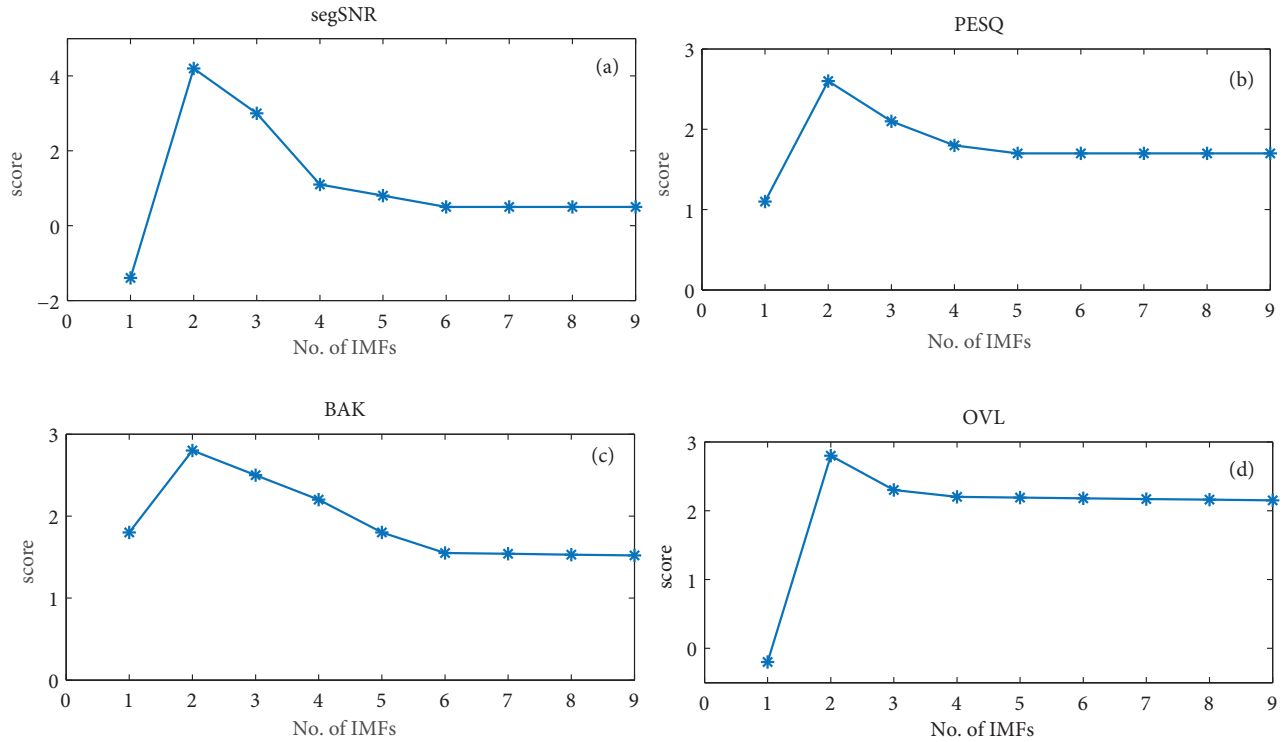


Figure 4. The objective quality measures improvement in the proposed method with respect to number of IMFs. (a)–(d) segSNR, PESQ, BAK, and OVL.

over time. To overcome these limitations first the noisy speech signal is decomposed into 2 IMFs after 15 iterations using EMD. The IMFs do not have abrupt power spectral changes over time. The obtained intrinsic mode functions are processed through NLM estimation with the input parameters bandwidth (λ), patch size (N), and neighborhood size (M) equal to 0.4σ , 8, and 80, respectively, where σ represents the standard deviation of the corresponding IMF. Finally, the enhanced speech signal is reconstructed from the processed IMFs. A clean speech signal taken from the TIMIT database, the signal after adding 0 dB of white noise from the Noisex-92 database, IMF-1, IMF-2, NLM estimation of IMF-1, NLM estimation of IMF-2, the enhanced speech signal, and corresponding spectrograms are shown in Figures 5a, 5b, 5c, 5d, 5e, 5f, 5g, and 5h–5n, respectively.

3. Speech enhancement baseline techniques

To evaluate the performance of the proposed EMD–NLM technique, three recently developed speech enhancement techniques are considered as baseline techniques in the present study.

FBE: This technique [17] is driven by the fact that the interfering source characteristics vary with time. In this work, the proposed formant-based foreground speech enhancement method contains the two modules, namely foreground speech segmentation and multistage foreground speech enhancement. The foreground speech is the speech signal recorded in a real-time scenario, whereas the background noise is the signal from other interfering sources. The foreground speech regions are firstly segmented from the background noise and later enhanced. Then the foreground speech is processed by LP analysis. Then alteration is performed on the regions around glottal closure instants in the LP residual signal and the formants to obtain the enhanced speech signal. In the FBE technique, musical noise is not introduced unlike other speech enhancement techniques such as the

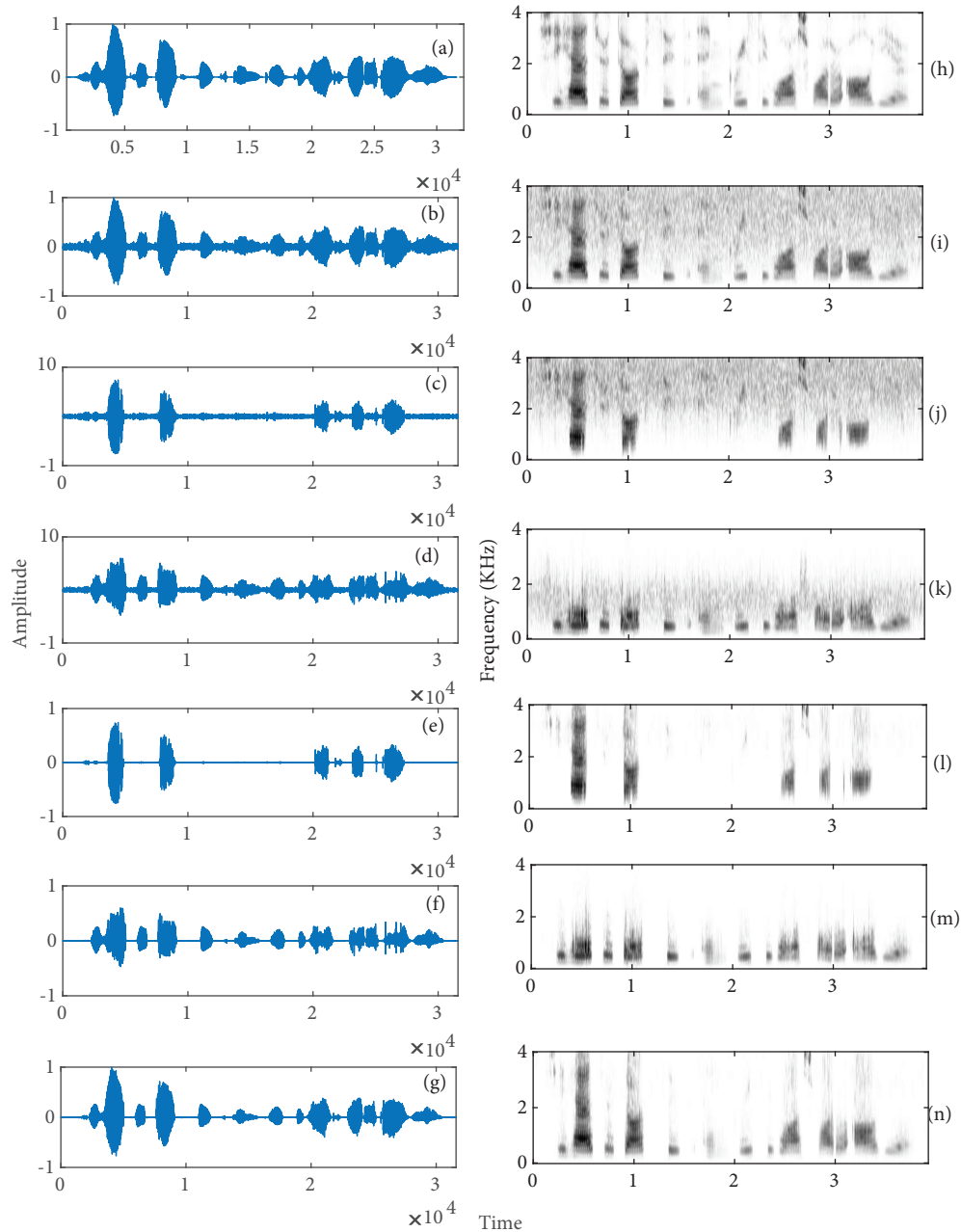


Figure 5. (a) A clean speech signal taken from the TIMIT database, (b) the signal after adding 0 dB of white noise from the Noisex-92 database, (c) IMF-1, (d) IMF-2, (e) NLM estimation of IMF-1, (f) NLM estimation of IMF-2, (g) enhanced speech signal obtained by using the proposed method, (h)–(n) spectrograms of clean, noisy, IMF-1, IMF-2, NLM estimation of IMF-1, NLM estimation of IMF-2, and enhanced speech signal, respectively.

minimum mean square error and spectral subtraction methods. The method proposed in [17] mainly focuses on enhancing the production and perceptual features of foreground speech rather than relying on modeling the different interfering sources. In this paper, this technique is named FBE.

EMD–VMD method: In this method [18], first empirical decomposition is used to decompose the noisy speech signal into oscillatory components called intrinsic mode functions (IMFs). Later, the Hurst exponent of

all obtained IMFs is computed to select the signal dominated IMFs. In this approach, the threshold value of the Hurst exponent is considered to be 0.5. Hence, $H \geq 0.5$ is obtrusive for the low-frequency noise components. Later VMD is applied to the summation of selected IMFs with input parameters $\alpha, \tau, S, tol, DC, and \omega_{init}$ to extract the narrowband components. Authors select only a few NCBs whose sum of their center frequency and standard deviation is less than or equal to 3600 Hz. All selected NCBs are processed through VMD with input parameters $\alpha, \tau, S, tol, DC, and \omega_{init}$ equal to 80, 0, 2, 10^{-7} , 1, and 0, respectively. The enhanced speech signal is obtained by adding all components except DC components. The improvement in the EMD–VMD method is essential to acquire better performance in terms of BAK and oSNR measures. Further, research is required to reduce the computational complexity of this method for real-time applications. This method is named EMD–VMD in our paper.

VMD–NLM: In this work [19], to enhance the speech signal, a two-level VMD–NLM-based method was utilized. First, the corrupted speech signal is decomposed into 12 VMFs by VMD algorithm. The data constraint fidelity balancing parameter is 320, timestep is 0, and tolerance of convergence is selected as 10^{-7} in this method. The 12 VMFs are clustered into four groups based on similarities in the location of their center frequency and magnitude spectrum. The NLM estimation is performed on each group with different value parameters to remove the noise components. All the parameter values are selected empirically. Finally, the NLM estimated signals are combined to get the enhanced signal. The method reported in [19] is named VMD–NLM in our paper.

4. Experimental results and discussion

The noisy speech signal is decomposed into 2 IMFs after 15 iterations using EMD. In order to suppress the ill effects of interfering noises, the obtained intrinsic mode functions are processed through NLM estimation with the input parameters bandwidth (λ), patch size (N), and neighborhood size (M) equal to 0.4σ , 8, and 80, respectively. Finally, the enhanced speech signal is reconstructed from the processed IMFs. The simulation result shows that the proposed method gives better performance in terms of subjective and objective quality measures.

4.1. Experimental dataset

To illustrate the efficiency of the proposed enhancement algorithm, we used a speech signal from the TIMIT database [20]. A set of 100 speech utterances from 50 male and 50 female speakers was used for experimental evaluations. The clean speech signal was corrupted by white noise, babble noise, and factory noise. The simulations were performed with three different SNR levels, 0, 5, and 10 dB. These noise sources were taken from the Noisex-92 database [21].

4.2. Performance evaluation

In order to assess the proposed method both subjective and objective measures are used.

Subjective assessment based on MOS score: To perform subjective analysis, we use mean opinion scores (MOS). An accurate subjective evaluation is a time-consuming task since it is challenging to find appropriate subjects for assessment evaluation. The parameters signal distortion (SIG), background noise (BAK), and overall quality (OVL) are weighed for subjective analysis. Thirty different subjects are chosen and provided with clean and enhanced speech files for evaluating the scores. The selected subjects involve 15 male and 15 female listeners having a mean age of 22 ± 4 years. All the listeners are well equipped with

English language skills and possess the capability to interpret the scores accurately. These listeners are pursuing postgraduate research in the signal processing domain and none of them is challenged with hearing inabilities. The evaluation was conducted in a silent atmosphere where there was no intrusion of any kind, and authentic headphones were provided suggesting a subtle atmosphere for assessment. The listeners were provided with the following instructions for each of such parameters.

1. The emphasis is on speech signal alone and scored using a five-point scale of signal distortion (SIG), namely [1 -Very Unnatural, 2 - Fairly Unnatural, 3 - Somewhat Natural, 4 - Fairly Natural, 5 - Very Natural].
2. The emphasis is on background noise (BAK) regions alone in terms of lesser distortion on a scale suggesting [1 - Very Intrusive, 2 - Somewhat Intrusive, 3 - Noticeable but not Intrusive, 4 - Somewhat Noticeable, 5 - Not Noticeable].
3. The emphasis is on the overall quality (OVL) with scales suggesting [1 - Bad, 2 - Poor, 3 - Fair, 4 - Good, 5 - Excellent].

The subjects mentioned above were equipped with three sets of similar files arranged in random order for assessment. Figure 6 represents the mean score value of the subjective assessment results for different methods from different subjects. The bar charts in Figure 6 show that the proposed method gives better performance in terms of mean opinion scores than the baseline techniques.

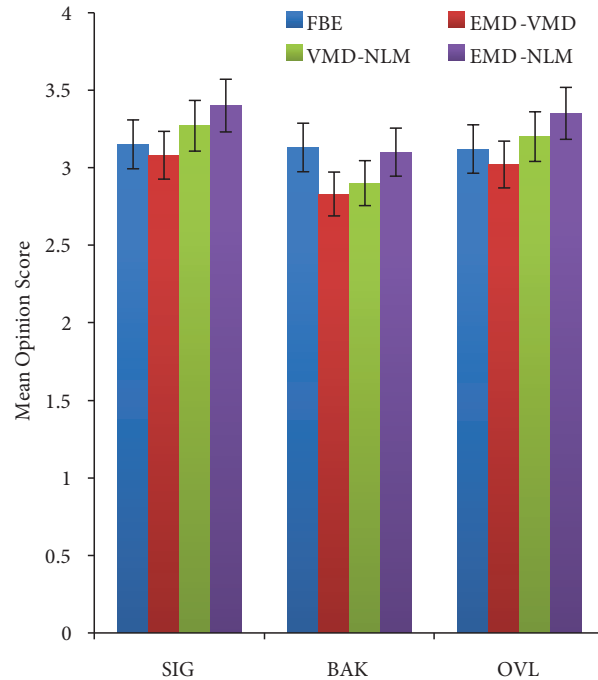


Figure 6. Bar graph representing the MOS obtained from different subjects, where SIG is signal distortion, BAK is background noise, and OVL is overall ratings.

Objective quality assessment: The perceptual evaluation of speech quality (PESQ) [22], scale of background intrusiveness (BAK) [22], scale of the mean opinion score (OVL) [22], and segmental signal to noise ratio (segSNR) [23] objective quality measures were used to evaluate the performance of the proposed algorithm. The perceptual evaluation of speech quality (PESQ) is one of the relevant objective measures that closely resemble subjective analysis. The results in terms of average values of SIG, BAK, OVL, and PESQ objective measures are shown in Table 1. Consistent improvements are observed for all three different noises used in the present study. Moreover, it can be observed that the segSNR is improved more in the proposed method under white noise compared with the other methods. The performance evaluation of EMD–NLM and VMD–NLM in terms of average values of BAK, OVL, segSNR, and PESQ for different SNR levels is shown in Figures 7a, 7b, 7c, and 7d, respectively. Figure 7 shows that the objective measures in the proposed method are approximately the same as those of VMD–NLM. However, the proposed method has less computational complexity (number of additions, integrations, and iterations) than VMD–NLM in real time.

Table 1. A comparison of the proposed method and existing speech enhancement techniques in terms of the mean score of background intrusiveness (BAK), the scale of the mean opinion score (OVL), segmental signal to noise ratio (segSNR), and perceptual evaluation of speech quality (PESQ). The simulation results are obtained by corrupting the clean speech signal with white, factory, and babble noises at different SNRs.

Noise	Input SNR (dB)	BAK			OVL			PESQ			segSNR		
		FBE	EMD-VMD	Prop.	FBE	EMD-VMD	Prop.	FBE	EMD-VMD	Prop.	FBE	EMD-VMD	Prop.
White	10	2.57	3.23	3.57	3.05	3.19	3.58	2.56	2.71	3.12	4.58	5.81	10.47
	5	2.36	2.7	3.12	2.73	2.85	3.12	2.34	2.4	2.84	3.09	4.59	7.56
	0	2.07	2.23	2.77	2.33	2.14	2.79	2.03	2.19	2.54	2.18	2.66	4.18
Factory	10	2.37	2.73	2.79	2.72	2.75	2.8	2.45	2.57	2.59	4.3	4.34	4.57
	5	2.12	2.18	2.27	2.4	2.39	2.42	2.23	2.35	2.29	2.97	2.52	2.58
	0	1.83	1.68	1.85	2.16	2.05	2.21	2.02	1.98	2.21	-0.62	-0.77	-0.37
Babble	10	2.21	2.79	2.85	2.55	2.61	2.67	2.36	2.44	2.48	4.54	4.42	4.63
	5	1.93	2.24	2.29	2.19	2.52	2.62	2.17	2.02	2.11	2.64	2.66	2.71
	0	1.61	1.72	1.74	1.81	1.96	1.21	1.79	1.85	1.84	-0.86	-1.01	-0.48

4.3. Computational complexity

To know the computational complexity of the contributed method, the program MATLAB (R2015a version) of the EMD–NLM method was executed on a computer with Intel Core i7 processor with a clock frequency of 3.10 GHz, along with the other developed speech enhancement techniques. The run time of other existed methods is normalized with respect to the run time of the EMD–NLM technique as shown in Table 2. The proposed technique showed that it takes minimal execution time when compared with other baseline techniques.

Table 2. Normalized mean processing time.

FBE	EMD–VMD	VMD–NLM	Proposed method
1.48	1.97	1.64	1

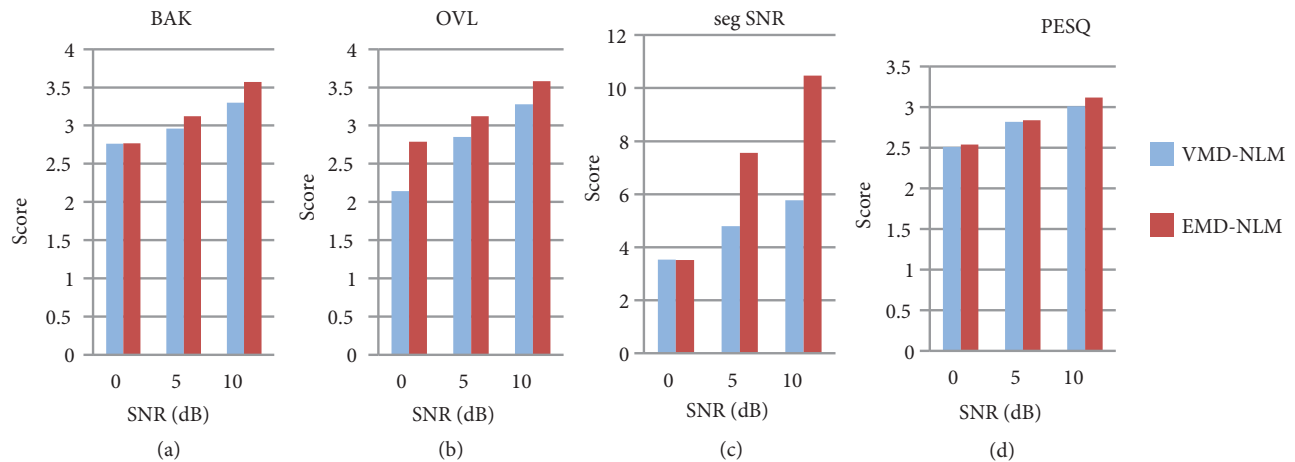


Figure 7. Performance evaluation of EMD–NLM and VMD–NLM in terms of average value of (a) BAK, (b) OVL, (c) segSNR, and (d) PESQ for different SNRs.

5. Conclusion

In this paper, we have proposed a novel speech enhancement method based on the combination of EMD and NLM whereby exploring the efficacy of both these methods. In this method, the noisy speech signal is first decomposed into two intrinsic mode functions by using the EMD algorithm. To suppress the ill effects of interfering noises, the obtained IMFs are processed through NLM estimation for better speech enhancement based on nonlocal similarities present in each IMF. The proposed method gives better performance when compared to the three recently developed methods, FBE, EMD–VMD, and VMD–NLM, in terms of the average value of SIG, OVL, BAK, segSNR, and mean opinion score. Simulation results show that the proposed method increases segSNR compared with other explored methods under white noise. The proposed method takes minimal execution time when compared with other baseline techniques. To evaluate the performance of the proposed system we use three different noises at different SNRs.

Acknowledgment:

The authors gratefully acknowledge the insightful comments of the editor and the anonymous reviewers.

References

- [1] Loizou PC. *Speech Enhancement: Theory and Practice*. 2nd Edition. Boca Raton, FL, USA: CRC Press, Inc., 2013.
- [2] Weiss MR, Aschkenasy E, Parsons TW. *Study and Development of the Intel Technique for Improving Speech Intelligibility*. Northvale, NJ, USA: Nicolet Scientific Corporation, 1975.
- [3] Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing* 2003; 11 (5): 466-475.
- [4] Lu Y, Loizou PC. Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty. *IEEE Transactions on Audio, Speech, and Language Processing* 2011; 19 (5): 1123-1137.
- [5] Virag N. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing* 1999; 7 (2): 126-137.
- [6] Lu Y, Loizou PC. A geometric approach to spectral subtraction. *Speech Communication* 2008; 50 (6): 453-466.

- [7] Proakis JG, Ling F, Nikias C. *Advanced Topics in Digital Signal Processing*. Minneapolis, MN, USA: Prentice Hall Professional Technical Reference, 1992.
- [8] Vaseghi SV. *Advanced Digital Signal Processing and Noise Reduction*. Hoboken, NJ, USA: Wiley & Sons, Inc., 2008.
- [9] Shao Y, Chang CH. A versatile speech enhancement system based on perceptual wavelet denoising. In: 2005 IEEE International Symposium on Circuits and Systems; Kobe, Japan; 2005. pp. 864-867.
- [10] Daqrouq K, Ibrahim IN, Abu-Isbeih, Daoud O, Khalaf E. An investigation of speech enhancement using wavelet filtering method. *International Journal of Speech Technology* 2010; 13 (2): 101-115.
- [11] Huang NE, Shen Z, Long SR, Wu MC, Shih HH et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: *Proceedings of the Royal Society of London, Series A: mathematical, physical and engineering sciences* 1998; 454 (1971): 903-995.
- [12] Khaldi K, Boudraa AO, Bouchikhi A, Alouane MTH. Speech enhancement via EMD. *EURASIP Journal on Advances in Signal Processing* 2008; 2008 (1): 873204. doi: 10.1155/2008/873204
- [13] Tracey BH, Miller EL. Nonlocal means denoising of ECG signals. *IEEE Transactions on Biomedical Engineering* 2012; 59 (9): 2383-2386.
- [14] Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. *SIAM Journal on Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal* 2005; 4 (2): 490-530.
- [15] Singh P, Pradhan G, Shahnawazuddin S. Denoising of ECG signal by non-local estimation of approximation coefficients in DWT. *Biocybernetics and Biomedical Engineering* 2017; 37 (3): 599-610.
- [16] Van De Ville D, Kocher M. SURE-based non-local means. *IEEE Signal Processing Letters* 2009; 16 (11): 973-976.
- [17] Deepak K, Prasanna S. Foreground speech segmentation and enhancement using glottal closure instants and mel cepstral coefficients. *IEEE/ACM Transactions on Audio, Speech and Language Processing* 2016; 24 (7): 1204-1218.
- [18] Upadhyay A, Pachori R. Speech enhancement based on MEMD-VMD method. *Electronics Letters* 2017; 53(7): 502-504.
- [19] Srinivas N, Pradhan G, Shahnawazuddin S. Enhancement of noisy speech signal by non-local means estimation of variational mode functions. In: 2018 Interspeech, Hyderabad, India; 2018. pp. 1156-1160.
- [20] Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS. DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. Nist Speech Disc 1-1.1. NASA STI/Recon Technical Report N, 1993.
- [21] Varga A, Steeneken HJ. Assessment for automatic speech recognition:II. Noisex-92: a database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication* 1993; 12 (3): 247-251.
- [22] Hu Y, Loizou PC. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing* 2008; 16 (1): 229-238.
- [23] Hu Y, Loizou PC. Evaluation of objective measures for speech enhancement. In: 2006 Ninth International Conference on Spoken Language Processing; Pittsburgh, PA, USA; 2006. pp.1-4.