Application of Variational Mode Decomposition on Speech Enhancement

Rashmirekha Ram
Department of Electronics and
Communication Engineering,
Siksha 'O' Anusandhan University
Bhubaneswar, Odisha, India.
ram.rashmirekha14@gmail.com

Sabyasachi Patra
Department of Electronics and
Communication Engineering,
Siksha 'O' Anusandhan University
Bhubaneswar, Odisha, India.
sabyapatra123@gmail.com

Mihir Narayan Mohanty
Department of Electronics and
Communication Engineering,
Siksha 'O' Anusandhan University
Bhubaneswar, Odisha, India.
mihir.n.mohanty@gmail.com

Abstract— **Enhancement of speech signal and reduction of noise from speech is still a challenging task for researchers. Out of many methods signal decomposition method attracts a lot in recent years. Empirical Mode Decomposition (EMD) has been applied in many problems of decomposition. Recently Variational Mode Decomposition (VMD) is introduced as an alternative to it that can easily separate the signals of similar frequencies. This paper proposes the signal decomposition algorithm as VMD for denoising and enhancement of speech signal. VMD decomposes the recorded speech signal into several modes. Speech contaminated with different types of noise is adaptively decomposed into various components is said to be Intrinsic Mode Functions (IMFs) by sifting process as in Empirical Mode decomposition (EMD) method. Next to it the denoising technique is applied using VMD. Each of the decomposed modes is compact. The simulation result shows that the proposed method is well suited for the speech enhancement and removal of noise by restoring the original signal.**

Keywords—Variational Mode Decomposition; Empirical Mode Decomposition; Intrinsic Mode Functions; Speech Enhancement

I. Introduction

Analysis of signal is a vital part for industry, research and academia. It is the great challenge for researchers in current decade. Different types of analysis can be performed based on users requirement. Signal estimation as well as estimation of its parameter is able to identify and confirm the model to be used, where noise is mostly contaminated with it. For enhancement of signal, it is highly essential to either remove or reduce the noise from it. More often we encounter these difficulties due to nonlinearities in signal and due to its nonstationary nature. For such purpose spectral analysis is an alternative. But accuracy level has to be taken care. Fourier transform is a popular and efficient tool for this purpose along with wavelet transform. Sometimes these methods are effective for specific cares and specific signals. But always it does not work. Also, for different signals different types of analysis is required. Consideration of Fourier transform may not work properly to its limitations, stated as follows:

For the nonstationary type of signal it fails to prove its efficiency. Also it cannot offer the temporal and spatial information clearly. Though it can be solved using wavelet transform, still it has certain demerits. One such example of nonstationary signal is speech and is considered in this piece of work.

The empirical mode decomposition of signal is an adaptive method and a powerful substitution to the Fourier and wavelet transform. It was proposed by Huang et al. He has suggested the technique for nonlinear and nonstationary signals. After decomposition of the signals, it can be reconstructed as the sum of the components along with the amplitude and frequencies parameters [1]. It can be said that the multi resolution method to perform space-spatial frequency decomposition as time-frequency analysis [2]. It is done empirically by removal of enhancement of the signal and IMF. Next to it the signal can be saved and used. This technique is self- adaptive. It can be used for an effective way of analysis of the instantaneous frequency of signals. The hidden information and structures can be obtained for further work. It has not been used by the researchers in mathematics and engineering. Though EMD has used by many researchers, it cannot be most popular due to its demerits in terms accurate mathematical model. Also choice of interpolation, sensitivity to noise and sampling are the factors of demerits. Therefore the birth of VMD occurred by Dragomiretskiy and Zosso as an alternative of EMD and can overcome the demerits of EMD. Its basic principle is same as EMD, but the center frequency of the mode has to be found so that the bandlimited modes can represent the original signal. This is explored in this work and applied for speech enhancement, though some others have used the same for different applications like classification and detection. To improve the quality of speech for hearing aid is highly essential and then the process is said to be the speech enhancement process. It occupies a great challenge for speech and signal processing researchers. Due to this process also further work can be performed such as classification, recognition, detection, coding, mobile communication and machine based approach. Many researchers have applied different types of algorithm for enhancement purpose. Speech enhancement is an active area of research in recent years. The algorithms used for this purpose may be listed as follows: spectral subtraction, statistical-model [3]. Estimation of noise spectrum is subtracted from spectrum of speech to estimation of clean speech spectrum [4-5]. But this method has drawbacks in case of musical noise. In case of

subspace or statistical – model based technique, decomposition of noise with noise signal vector space into speech subspace and noise subspace. Then the resultant speech is enhanced by projecting it into the speech subspace [6]. Also different adaptive algorithms have been used for enhancement application [7-8].

Speech quality might be degraded in the presence of background noise. To improve the voice quality, it is essential to include the Speech enhancement (SE) technique for different applications. Various methods have been studied in the literature, such as Spectral subtraction (SS), Wiener filtering (WF), Kalman filtering, and model-based methods. Similarly multiband spectral subtraction method for speech enhancement has been proposed in [9]. In that case authors are considered the colored noise contaminated with speech. Another filter named as Ephraim Malah filtering has been proposed in [10]. HMM based method also has been applied for speech enhancement in [11]. Though different methods as given in literatures along with wavelet and DCT domain have been applied by researchers the proposed method has the significance to apply in this work.

There are many applications in speech enhancement based on adaptive filtering algorithms. Many methods have been used by different researchers from several decades. Some of them are summarized below. An adaptive β-order generalized spectral subtraction method was proposed by Junfeng Li *et.al* [12]. The characteristics of the speech signal changes very rapidly. So the adaptive algorithms are required to reduce the noise or enhance the signal quality. To reduce noise from the speech signals, Sayed. A. Hadei and M. Lotfizad have used fast affine projection algorithm and fast Euclidean direction search algorithm [13]. To reduce the matrix operations used in general Kalman filtering, in [14] a fast adaptive Kalman filtering was used for enhancing the speech signal. The calculating time was reduced in the proposed algorithm. The Convex Combination of WSLMS (CC-WSLMS) algorithm was proposed to improve the adaptive filter performance. According to the variance of the output power, the filter weight vectors were modified [15].

Speech is one type of biomedical signal and its nature is nonstationary. For different biomedical signals, authors have approached for detection and classification with many methods as in [16-17]. EMD method is also used for feature extraction of EEG signal. Neural network with fuzzy algorithm is also used for testing the performance of different features. An incremental feature analysis technique is also used for feature selection and classification [18-19].

This paper is organized as follows: Section I introduced the work along with some related literature. Method for enhancement is explained in section II. It follows the result in section III and section IV concludes the work.

II. VMD ALGORITHM FOR SPEECH ENHANCEMENT

Different methods have been used for speech enhancement as described in literature in section I. Mode decomposition is one of the efficient methods and has been used by the name

of EMD, EWD (Empirical Wavelet Decomposition) and VMD. But these are not used for speech enhancement. Like EMD, VMD helps in decomposition of a signal resulting different IMFs (Intrinsic Mode Functions) which are basically depends on the number of level taken into consideration. Further out of these modes decomposition algorithms, VMD found better than other two for which authors in this work are applied the same for speech enhancement.

Empirical Mode Decomposition (EMD) and Empirical Wavelet Transform (EWT) methods are widely used for decomposition of signal and image. These methods have some drawbacks like other decomposition technique. Basically major disadvantage of EMD is that it is mode mixing and being affected by noise. Wavelet filters are used in EWT and from that it is easy to extract IMFs. Here, filters are used for perfect reconstruction of signal. VMD is an adaptive method and helps in decomposition of the signal into N number of intrinsic mode functions. The modes are extracted concurrently in VMD technique. It is a non-recursive mode decomposition technique. The ensemble band-limited intrinsic modes of center frequencies reproduce the input signal. The following steps are involved in speech enhancement using VMD method.
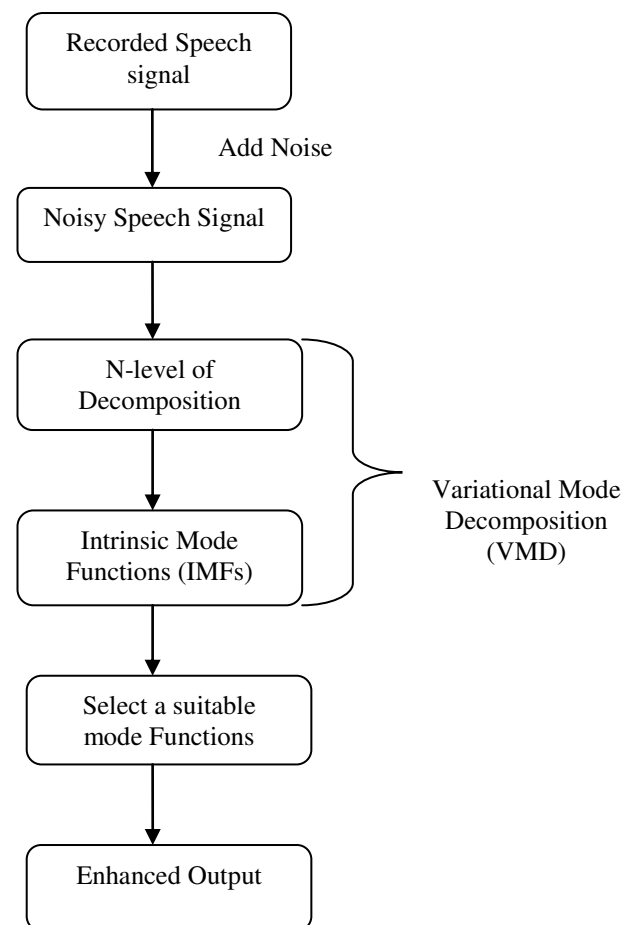

Fig.1. Speech Enhancement using Variational Mode Decomposition method

Each mode is assumed as seeking k modes i.e. $m_{k..}$ Minimizing the estimated bandwidth of each mode, and its constraint is sum of each mode equal to the input signal. Now configuration steps are as follows:

(1) To obtain the single frequency spectrum for each mode m_k, the analytic signal is computed by using the Hilbert transform as follows:

$$(\delta(t) + \frac{j}{\pi t}) * m_k(t) \quad (1)$$

(2) The frequency spectrum is shifted to 'baseband' for each m_k by combining with exponential component $e^{-j\omega_k t}$ as:

$$[(\delta(t) + \frac{j}{\pi t}) * m_k(t)] \, e^{-j\omega_k t} \quad (2)$$

(3) Squared L^2-norm of the gradient of the demodulated signal can be computed as:

$$\min_{\{m_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * m_k(t) \right] e^{-j\omega t} \right\|^2 \right\} \quad (3)$$

$$\text{Where} \sum_k m_k = f$$

$$\{m_k\} = \{m_1,...., m_k\}$$

$$\text{And} \ \{\omega_k\} = \{\omega_1,....,\omega_k\}$$

In this work, VMD is applied for speech enhancement. The voice signal affected by noise is decomposed into three modes giving rise to an enhanced signal after third level of decomposition. So VMD is an effective method for decomposition of a voice signal. VMD process basically includes three important concepts i.e Wiener filter, Hilbert transform and frequency mixing.

Main aim of VMD is to decompose the input signal into different modes of IMFs. Voice signal for "Hello, one one one", is added with noise and then that noisy signal of voice "Hello, one one one", decomposed into three levels.

III. RESULTS AND DISCUSSION

Initially the voice signal is collected in a closed room. The signal is partially noisy and shown in Fig. 2. A random noise of variance 0.1 is added to the recorded signal for the test purpose and is shown in Fig. 3 as the noisy speech signal. The utterance is shown for "Hello, one one one". Different types of noisy signal have been generated and also collected from [7]. The noisy voice signal undergoes variational mode decomposition to produce three decomposed voice signal as the number of modes initialized as K=3(Where, K is the number of modes). First level of decomposition provides the noise in the signal which is more as compared to second level of decomposition as shown in Fig. 4(top two signals). But third level of decomposition which separates the noise resulting in enhanced signal as in Fig 4 (bottom one). The level of noise reduction for different noisy speech signals are measured in terms of SNR.
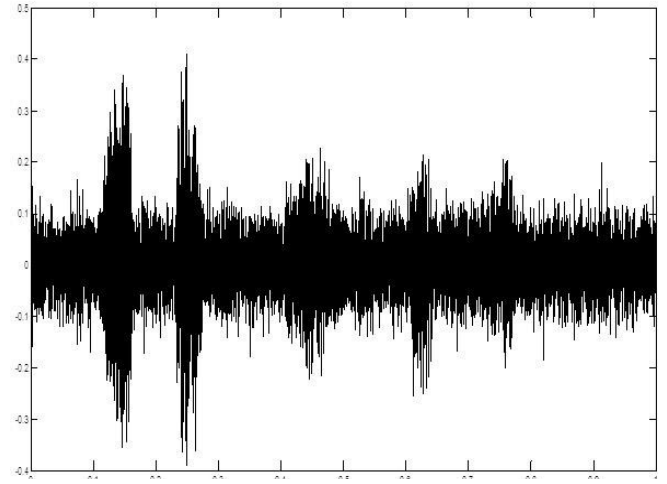

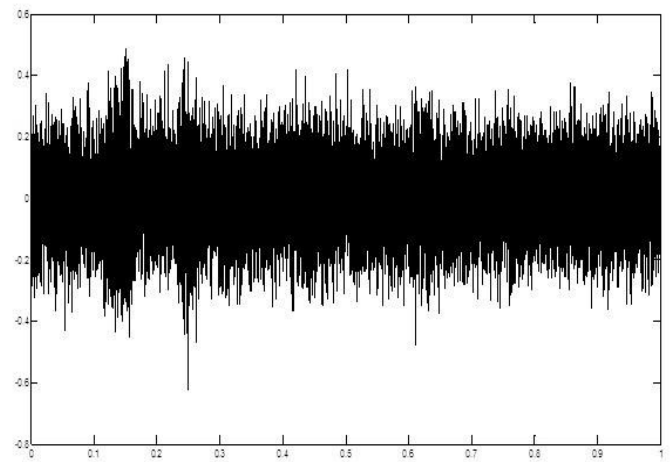Fig. 2. Voice signal for "Hello, one one one"

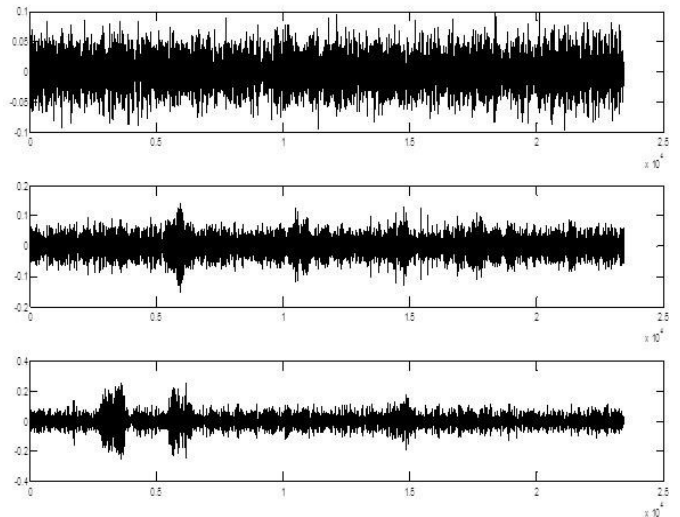Fig. 3. Noisy signal of "Hello, one one one"

Fig. 4. Decomposed enhanced signal

Signal to noise ratio is an important criterion for determining the quality of Speech. It is represented as

$$(SNR)_{dB} = 10\log_{10}\left(\frac{S_0}{N_0}\right) \qquad (4)$$

Where S_0 is the total signal energy and N_0 is the total noise energy.

Due to lack of space, all the experimental voices are not given in figure. For sample, one of the figures is shown. For other voices the tables are given as follows. Table-I shows the SNR level for different voice samples contaminated with different types of noise. The SNR level before enhancement and after enhancement using proposed method is shown.

For the sake of comparison, the proposed method is compared with different methods approached earlier. It is shown in table II. The improvement of SNR is 23.7143 dB for VMD which is much greater than the other methods. The listening tests are also performed for different enhanced output signals. The more clear sound is obtained from VMD.

TABLE I. VOICE SIGNAL WITH DIFFERENT NOISE LEVEL ENHANCEMENT

Voice with different Noise	SNR before enhancement (dB)	SNR after enhancement (dB)
Random noise	11.2342	16.6292
Babble noise	10.0478	15.7152
Train noise	12.3608	17.2894
Airport noise	11.7143	17.0376

TABLE II. COMPARISON OF PROPOSED METHOD

Methods	SNR Improvement (dB)
Spectral Subtraction (SS) [7]	4.1407
Least Mean Square (LMS) [7]	6.5245
Recursive Least Square(RLS) [7]	14.3883
VMD	23.7143

IV. CONCLUSION

This paper proposes the speech enhancement scheme. The proposed approach for enhancement has been applied successfully. At the time of training the noisy speech has been applied directly, the effects of this algorithm comprises of different steps to enhance a noisy speech signal and have been demonstrated in the result section. Results obtained for speech signal shows that the enhancement using VMD method is more effective than the other methods. The proposed method is a promising method in the area of signal and speech enhancement and can be used further. As the Wiener filter is included in VMD, so least square method is more suitable than the average mode based technique and the maximum average technique. These may be analyzed and can be kept for future work. For better signal reconstruction the effect of sampling as well as interpolation technique has to be taken care and kept for a future work.

REFERENCES

[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Royal Soc. A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[2] Mihir Narayan Mohanty, Aurobinda Routray, Ashok Kumar Pradhan, P.Kabisatpathy, "Power Quality Disturbances Classification using Support Vector Machines with optimized time-frequency kernels", International Journal of Power Electronics,Vol.4, No 2, pp181-196, 2012.

[3] P. Loizou. Speech Enhancement: Theory and Practice. CRC Press, 2007.

[4] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on Acoustics, Speech and Signal Processing, 27(2):113 – 120, Apr. 1979.

[5] Yang Lu and Philipos C. Loizou, "A geometric approach to spectral subtraction. Speech Communication", 50(6), pp 453–466, 2008.

[6] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement. Speech and Audio Processing", IEEE Transactions on, 3(4), pp 251 –266, Jul. 1995.

[7] Rashmirekha Ram, Mihir Narayan Mohanty, "Performance Analysis of Adaptive Algorithms for Speech Enhancement Applications", Indian Journal of Science & Technology, 2016. *(In press)*

[8] V. Anoop, P.V.Rao and M. Nidhin, "Performance Analysis of Speech Enhancement Methods Using Adaptive Algorithms and Optimization Techniques", IEEE, ICCSP,2015.

[9] S.Kamath, P.Loizou, "A Multi-band spectral subtraction method for enhancing speech corrupted by colored noise", International Conference on Acoustics, Speech and signal processing, Vol.4,pp 4160-4164, 2002.

[10] Y. Ephraim, D. Malah, "Speech Enhancement using a Minimum Mean Square Error Log Spectral Amplitude Estimator", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol-33, pp 443-445, 1985.

[11] H.Sameti, "HMM-based strategies for Enhancement of Speech Signals Embedded in Nonstationary noise", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol-6, pp445-455, 1998.

[12] Junfeng Li , Shuichi Sakamoto, Satoshi Hongo, Masato Akagi, Yoiti Suzuki, "Adaptive β-order generalized spectral subtraction for speech enhancement", Signal Processing, Elsevier Journal, No-88, pp 2764–2776, 2008.

[13] Sayed. A. Hadei, M. Lotfizad, "A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement", International Journal of Computer and Electrical Engineering, Vol. 2, No. 2, 2010.

[14] Quanshen Mai, Dongzhi He, Yibin Hou, Zhangqin Huang, "A Fast Adaptive Kalman Filtering Algorithm for Speech Enhancement", IEEE International Conference on Automation Science and Engineering Trieste, Italy, August 24-27, 2011.

[15] Wu Caiyun, "Study of Speech Enhancement System Using Combinational Adaptive Filtering", International Conference on Computer Distributed Control and Intelligent Enviromental Monitoring, 2012.

[16] Semwal, Vijay Bhaskar, Kaushik Mondal, and G. C. Nandi. "Robust and Accurate Feature Selection for Humanoid Push Recovery and Classification: Deep Learning Approach." Neural Computing and Applications, pp 1-10, 2015.

[17] Kumari, Pinki, and Abhishek Vaish. "Feature-Level Fusion of Mental Task's Brain Signal for an Efficient Identification System." Neural Computing and Applications, 27.3, pp 659-669, 2016.

[18] Kumari, Pinki, and Abhishek Vaish. "Information-Theoretic Measures on Intrinsic Mode Function for the Individual Identification using EEG sensors." IEEE Sensors Journal 15.9 pp 4950-4960, 2015.

[19] Semwal, V.B., Singha, J., Sharma, P. et al., "An Optimized Feature Selection Technique based on Incremental Feature Analysis for Bio-Metric Gait Data Classification" Multimed Tools Appl., DOI:10.1007/s11042-016-4110, 2016.