

# Speech Enhancement in Wavelet Domain using Principle Component Analysis and Enhancement Filters

B. Kirubagari

Assistant Professor

Department of Computer Science and Engineering,  
Annamalai University, Chidambaram, India

S. Palanivel

Professor

Department of Computer Science and Engineering,  
Annamalai University, Chidambaram, India

## ABSTRACT

The aim of speech enhancement is to improve the perceptual quality and intelligibility of the speech by reducing the background noise. This paper proposes a technique in wavelet domain to enhance the signal. The signal is decomposed into approximation coefficients and detail coefficients which are filtered separately using spectral subtraction and Wiener filter. The signal is reconstructed by transforming it into time domain by applying inverse wavelet domain. Wavelet features of the noisy speech signal are extracted and the dimension of the features is reduced using Principle component analysis (PCA). Experiments are conducted on noisy speech signal database (NOIZEUS), which consists of speech signals corrupted by eight different real world noises recorded at different signal-to-noise (SNR) levels. The performance of the proposed algorithm is evaluated using SNR, which is a standard measure of the amount of background noise present in a speech signal and Mean opinion score (MOS). Experiments results show the increase in the efficiency of the proposed enhancement algorithm.

## Keywords

Speech enhancement, PCA, Wavelet domain, SNR

## 1. INTRODUCTION

Speech constitutes a fundamental activity for the interaction of human being with the surrounding world. Speech is the vocalized form of human communication. It is based up on the syntactic combination of lexical and names and also ten times faster than written communication to transfer information. And it is rich too. The objective of speech enhancement is to improve the quality and intelligibility of the speech. The main goal is to reduce the noise in the speech. Due to the practical importance, noise robustness has become an active research area in various speech processing techniques.

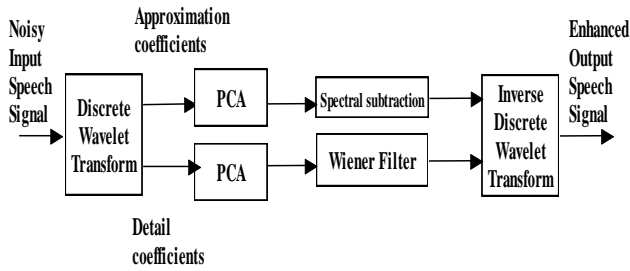
Speech constitutes a fundamental activity for the interaction of human being with the surrounding world. The main purpose of speech enhancement is to improve the perceived quality of speech presented to listener. The presence of background noise causes the degraded quality or intelligibility. Noisy environment also reduces listeners' ability to understand what is spoken. The corrupted speech consists of the speaker voice, background noise and computing noise generated by the device. When the speech is transmitted through some medium, some transmission noise also gets mixed with the speech. Speech enhancement techniques have been applied to improve the quality and intelligibility of the noise corrupted speech by using digital processing tools [1].

Noise reduction techniques started up by Schroeder who proposed an analog implementation of the spectral magnitude method. Boll [2] reinvented this method in digital domain. Spectral subtraction is a popular method to eliminate stationary additive background noise. Ephraim and Malah proposed a minimum mean-squared error (MMSE) amplitude estimator in the log-spectral domain. This estimator is more consistent with human auditory perception [4]. K.K.Paliwal and A.Basu has tried with kalman filter on AR coordinates [5].

The Wiener filter is the most fundamental approach. This is an optimal filter which minimizes the Mean Square Error (MSE) between the desired signal and the estimated signal [6]. A new speech enhancement approach for musical noise reduction is proposed by combining spectral subtraction and the conventional Wiener filtering in series connection to construct a two-stage hybrid system in frequency domain [12].

Another important approach proposed by Ephraim and Van trees is based on signal subspace decomposition [3]. The idea of this algorithm is based on the fact that the noisy signal can be decomposed into a signal plus noise subspace and an orthogonal noise subspace, which will be removed first. Processing is performed only on the vectors in the signal plus noise subspace. Decomposition of the speech signal is performed by applying Eigen value decomposition (EVD) or singular value decomposition (SVD) or by applying the Karhunen-Loeve transforms (KLT) [4].

In order to improve the robustness and efficiency of the speech enhancement, it is feasible to combine two or three algorithms in different manners or in different domains. Wavelet transform (WT) has recently been evolved as a powerful tool for enhancement which helps in removing noise from speech and image signals effectively [13]. The localizing property of the wavelets is very helpful in thresholding and shrinking the wavelet coefficients which helps in separating the signal from noise. The signal is transformed into wavelet domain by applying wavelet transform. The proposed method of speech enhancement is shown in Fig.1.



**Fig.1. Proposed Method of speech enhancement**

This paper is organized as follows: Section 2 depicts the basics of Wavelets. Section 3 explains about PCA. Section 4 describes spectral subtraction method. Section 5 explains the wiener filters. Section 6 gives the proposed methodology and in Section 7 the experimental results were given. The speech quality evaluation is presented in Section 8 and the conclusion is given in Section 9.

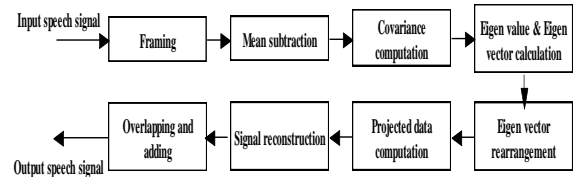
## 2. Wavelets

Wavelet Transform is a very useful in achieving data compression, detecting features and de-noising signals. Wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet. There are different kinds of wavelets which include smooth wavelets, compactly supported wavelets, wavelets with simple mathematical expressions, and wavelets with simple associated filters, etc. The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two- the so called dyadic scales and positions. DWT breaks the signal into low frequency components known as the approximation coefficients and high frequency components known as the detail coefficients.

## 3. PCA:

PCA is a powerful and widely used tool for analyzing data and dimension reduction. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Another advantage of PCA is that once these patterns are found, the data can be compressed, i.e. by reducing the number of dimensions, without much loss of information [14]. PCA consists of computation of the eigenvalue decomposition or singular value decomposition of a data set, usually after mean centering the data for each attribute [15, 16].

The speech signal is framed into short frames to exploit the slowly varying time properties. Thus the data is arranged in two dimensions. The empirical mean is subtracted from each column of matrix. If error is to be minimized then the covariance matrix has to be maximized. The covariance matrix can be guaranteed to have real Eigen-values. These Eigen-values may be sorted in descending order and the associated Eigen-vectors taken as the. basis vectors that provide the maximum information to reconstruct the signal. In the data approximation, dimensions corresponding to the smallest Eigen-values are omitted. The signal is overlapped and added to get the original signal as shown in Fig.2



**Fig.2. Flowchart of PCA**

## 4. Spectral Subtraction

The speech enhancement algorithm aims to improve the perceptual aspects of speech such as overall quality or intelligibility. Spectral subtraction is a process of subtracting estimated noise spectrum from the noisy signal. It acts as a good pre-processor for other speech processing techniques.

The basic idea of spectral subtraction is performed in the spectral domain by operating on the Fourier transformation of the observed samples. The estimation of the noise-reduced speech spectrum is obtained from the spectral magnitude of the noisy speech signal by subtracting an estimated mean spectral magnitude of the noise [9]. Spectral subtraction is mainly suitable for stationary or very slow varying noises. The main weakness of the spectral subtraction is the production of an annoying noise called musical noise[2], which is suffered from the over-subtracting of the spectral proposed by Boll in 1978[6].

## 5. Wiener Filter

The Wiener filter is a popular technique that has been used in many signal enhancement methods [8]. The basic principle of the Wiener filter is to obtain a clean signal from that corrupted by additive noise. It is required to estimate an optimal filter for the noisy input speech by minimizing the mean square error between the desired signal and the estimated signal [8, 10]. The conventional wiener filters have the characteristics of suppressing the noise frequencies with the other speech frequencies unchanging [11]. The Wiener filter is based on the ensemble average spectra of the signal and noise. The Wiener filter rule is derived from the optimal filter theory.

## 6. The Proposed Method

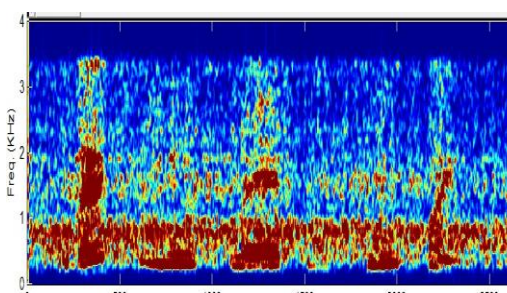
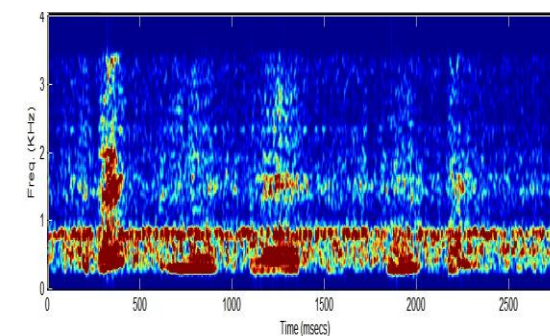
The proposed system is illustrated by Fig.1. It is a three stage approach which is used to enhance the noisy speech. The wavelet coefficients are obtained by applying discrete wavelet transform on the input noisy signal. The dimensions of the approximate coefficients and detail coefficients are reduced using PCA. The signal is decomposed based on Subspace decomposition using the Eigen values. The dimension reduced approximation coefficients are filtered using the spectral subtraction which suppresses the stationary noise components. The dimension reduced detail coefficients are filtered using the Wiener filter which is employed to reduce the real

background noise. Principles of the signal processing approaches illustrate the feasibility of the proposed method. The signal is reconstructed after transforming it into time domain.

## 7. Experimental setup

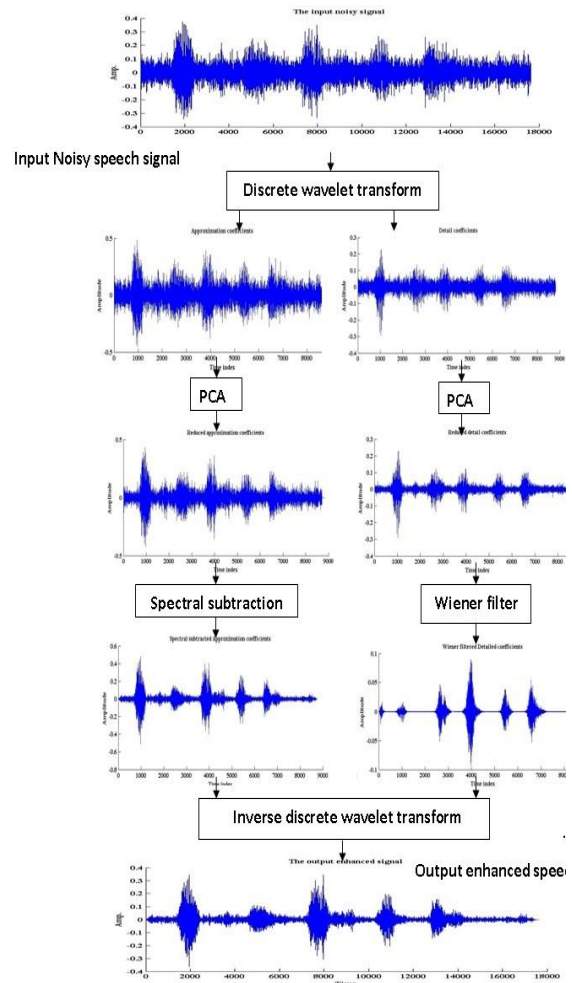
Experiments are done on noisy speech signal database (NOIZEUS) which is corrupted by eight different noises at 0dB, 5dB, 10dB and 15dB SNR levels. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, and airport and railway-station noise [17]. The speech data in time domain is transformed into wavelet domain for easy removal of noise. The wavelet coefficients are obtained by applying discrete wavelet transform and thus extracting the approximate and detail coefficients. The dimensions of the extracted coefficients are reduced by applying PCA. The number of Eigen values chosen forms the basis vectors of the signal subspace. The noisy approximation coefficients which contain low frequency components are then filtered out using spectral subtraction which again removes stationary background noises. The minor information stored in the detailed coefficients is also filtered using wiener filter. The signal is then retransformed into its original form by applying its inverse transform. Thus the noise is removed from the signal in the proposed method.

The number of Eigen values chosen is used for the reconstruction of the signal. The signal will be distorted if less number of Eigen values is chosen. The noise will not be eliminated if more number of Eigen values is chosen. Fig. 3.a. and Fig.3.b. shows the spectrographs of the reconstructed signals using different Eigen values. It



**Fig.3.a** Reconstructed signal of signal corrupted by car noise at 0db SNR using 20 Eigen values.  
**b.** Reconstructed signal using 40 Eigen values.

Fig.4 shows the methodology and results of the proposed work. The signal coefficients are transformed into wavelet domain which separates the high frequency components and low frequency components. The figure clearly shows that the subspace method using PCA separates the speech and noise space. The dimensions of the feature set of the signal are reduced and thus some noise in signal is eliminated. The spectral subtraction and wiener filter filters the noise components present in the signal coefficients separately. The enhanced signal is reconstructed using inverse discrete wavelet transform.



**Fig.4. Methodology and results of the proposed method**

## 8. Performance analysis

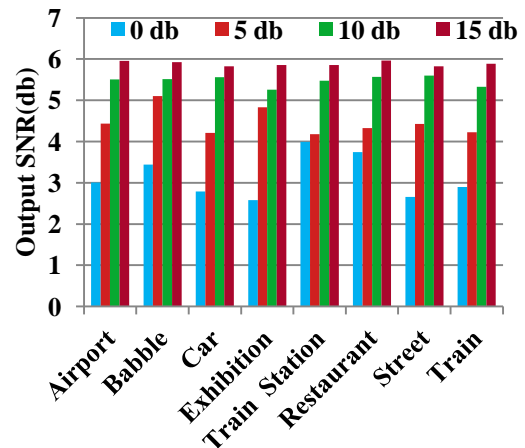
The performance of this algorithm is evaluated by using SNR and MOS. SNR is defined as the ratio of signal power to the noise power. It is an important time domain feature in determining the quality of speech [6].

**Table1. Output SNR of Enhanced speech for all sorts of noises**

Noise	SNR(db)			
	0db	5db	10db	15db
Airport	3.0040	4.4366	5.5040	5.9545

Noise		SUBJECTIVE SPEECH INTELLIGIBILITY TEST		
		Male Subject opinion score1	Female subject opinion score	Mean opinion score
Airport	0 db	3	3	3
	5 db	4	3	3.5
	10 db	4	5	4.5
	15 db	4	5	4.5
Babble	0 db	3	3	3
	5 db	4	5	4.5
	10 db	5	5	5
	15 db	5	5	5
Car	0 db	2	2	2
	5 db	3	4	3.5
	10 db	5	4	4.5
	15 db	5	5	5
Exhibition	0 db	1	2	1.5
	5 db	2	2	2
	10 db	3	2	3.5
	15 db	4	3	3.5
Restaurant	0 db	2	1	1.5
	5 db	4	4	4
	10 db	5	4	4.5
	15 db	5	5	5
Street	0 db	2	2	2
	5 db	2	3	2.5
	10 db	3	4	3.5
	15 db	4	5	4.5
Train	0 db	2	3	2.5
	5 db	3	3	3
	10 db	4	4	4
	15 db	5	5	5
Railway Station	0 db	1	1	1
	5 db	2	2	2
	10 db	3	4	3.5
	15 db	4	4	4

Babble	3.4445	5.1024	5.5128	5.9288
Car	2.7863	4.2141	5.5599	5.8248
Exhibition	2.5789	4.8284	5.2617	5.8528
Train Station	3.9903	4.1757	5.4738	5.8586
Restaurant	3.7480	4.3286	5.5654	5.9674
Street	2.6566	4.4311	5.6029	5.8227
Train	2.8981	4.2255	5.3245	5.8838



**Fig.5. SNR of Enhanced speech for all sorts of noises**

High SNR values show that the output enhanced signal is preserving the signal content as possible. The proposed approach gives better SNR for Airport noise, Babble noise, Restaurant noise and Train station as shown in Fig.5. Restaurant noise gives better performance than street noise though they look similar. Street noise and Train noise gives better performance in high SNR values than low SNR values. Exhibition and Train station noise gives a moderate performance as shown in Table.1.

Subjective criteria are measured by a group of listeners that give their personal evaluations of the encoder quality. MOS is the most widely used subjective measure of voice quality and is recommended by the International telecommunication Union (ITU). A MOS value is normally obtained as an average opinion of quality based on asking people to grade the quality of speech signals on a five-point scale

**Table.2. Opinion Score and Mean Opinion Score of Speech Intelligibility Test on Male and Female Subjects.**

(Excellent, Good, Fair, Poor, and Bad) under controlled conditions. Table.2 shows that the female speakers can recognize the reconstructed signal. Airport noise, Babble noise and Car noise can be recognized easily than the other noises. The exhibition noise gives poor values.

## 9. Conclusion:

This paper addresses the problem of noise reduction of noise in speech using the combination conventional spectral subtraction and Wiener filter in wavelet domain. Features of the noisy speech are extracted. The features are transformed into wavelet domain in which the features are separated as high frequency and the low frequency components. . Using the subspace technique, PCA the dimensions of the features are reduced which in turn eliminates some noise too. The reduced high frequency components are filtered using spectral subtraction and low frequency components are filtered using wiener filter. The SNR shows that this method gives improvement

## 10. REFERENCES

- [1] J. Benesty, S. Makino and J. Chen. 2005. Speech Enhancement. Springer, Berlin Germany.
- [2] S. F. Boll . 1979. Suppression of acoustic noise in speech using spectral subtraction . IEEE Trans. Acoustic., Speech. Signal Processing. 27:113–120.
- [3] Y. Ephraim and H. L. Van Trees.1995. A Signal Subspace Approach for Speech Enhancement. IEEE Trans. Speech Audio Processing. 3:251-266.
- [4] Anuradha R. Fukane, Shashikant and L. Sahare.2011. Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments. International Journal of Scientific & Engineering Research.2:5
- [5] A. Rezayee and S. Gazor.2001.An adaptive KLTapproach for speech enhancement. IEEE Trans.Speech Audio Processing.9:87-9.
- [6] M.A. Abd El- Fattah, M.I. Dessouky, S.M. Diab and F.E. AbdEl-samie.2010.Adaptive Wiener Filtering approach for Speech Enhancement. International Conference on Speech and Signal Processing.
- [7] Mingzhe Zhu, HongbingJi, Falong Luo and Wei Chen.2007.A robust speech enhancement scheme on the basis of bone-conductive microphones.IEEE.353-355.
- [8] Wang Guang-Yan, ZHAO Xiao-qun and WANG Xia.2009.Musical Noise Reduction Based on Spectral Subtraction combined with Wiener Filtering for Speech Communication. IEEE Transaction on signal processing.
- [9] ThomasEsch and Peter Vary.2009.Efficient musical noise suppression for speech systems. Acoustics, Speech and Signal Processing, ICASSP.4409 – 4412.
- [10] M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab, and F. E. Abd El-samie.2008.Speech enhancement using an adaptive wiener filtering approach. Progress In Electromagnetics Research M.4:167–184.
- [11] L.R.Rabiner, R.W.Schafer.2005.Digital Processing of Speech Signals. Pearson Education Pte. Ltd.Second Edition. New Jersey.
- [12]. Wang Guang-yan, Zhao Xiao-qun, and Wang Xia.2009. Musical Noise Reduction Based on Spectral Subtraction Combined with Wiener Filtering for Speech Communication.IET International Communication Conference on Wireless Mobile & Computing (CCWMC 2009).726–729.
- [13].Saeed Ayat, M.T. Manzuri-Shalmani, and RoohollahDianat.2006.An improved wavelet-based speech enhancement by using speech signal features. Speech Communication. 48:927– 940.
- [14] Smith, L.I. 2002.A tutorial on Principal Components Analysis. [http://csnet.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).
- [15] Jonathon Shlens .2005.A Tutorial on Principal Component Analysis. Systems Neurobiology Laboratory, Salk Institute for Biological Studies Jolla, CA92037 and Institute for Nonlinear Science, University of California, San Diego La Jolla, CA 92093-0402.
- [16] Qin, S. J.; Dunia R.1998. Determining the number of principal components for best reconstruction. In *IFAC DYCOPS'98*. Greece
- [17] Hu, Y. and Loizou, P.2007.Subjective evaluation and comparison of speech enhancement algorithms. Speech Communication.49:588-601.
- [18] ITU-T Rec. P.800 (1996). Methods for Subjective Determination of Transmission Quality.