IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 113-118
www.iosrjournals.org

Speech Enhancement by using Transform Domain Techniques

Mr. Pradnyesh Deshmukh[1], Mr. D. G. Bhalke[2]

[1] *Department of Electronics and Telecomm Engineering, JSPMs, `Rajarshi Shahu College of Engineering, Pune, India.*
2 Department of Electronics and Telecomm Engineering, JSPMs, `Rajarshi Shahu College of Engineering,

Abstract : *The techniques of speech enhancement are used to minimize the undesirable external noises; the speech signal get degraded due to these external noises and this noises are background noise, wind noise, and other noises. These noises generated due to environmental sources such as vehicular noise, street noise, babble noise, etc. Therefore, the speech enhancement plays an important role to reduce unwanted noise from the speech signal. This paper shows the single channel, the speech enhancement techniques based on a Spectral Subtraction Method (SSM) and Wavelet Packets Method (WPM) have been used for removing the noises. The performance of enhanced speech signal has been evaluated by measuring the speech quality parameters like as cross-correlation, average absolute distortion (AAD), Peak Signal to noise ratio (PSNR and Mean Square Error (MSE). The speech signal recorded in male voice and the Noizeus Speech database with different speech signals has been used to test the performance of this system. The experimental result shows that the Wavelet Packet method is more suitable for speech enhancement.*

Keywords **-** *Wavelet Packet Method (WPM), Signal to Noise Ratio (SNR), Minimum Mean Square Error (MMSE), Spectral Subtraction Method (SSM), Speech Enhancement, Average Absolute Distortion (AAD), Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE).*

I. INTRODUCTION

The usage of speech communication and the related products like voice over internet protocol (VoIP), teleconferencing, audio video calls, hearing aids, etc. has been increased due to the revolution in mobile phone technology and its different applications. Today, most of the speech communications applications are work fine in a noise-free environment where as the speech communication quality deteriorates in an external noisy condition. The speech enhancement method is used for the improvement of a quality of noisy speech signal so that there will be less noise should be present in the speech signal. A noise may be generated by different sources of environment conditions such as vehicular noise, street noise, babble noise, exhibition noise, etc. It is very difficult to estimate various types of noises and their time variations in speech communication systems. Therefore, the complete noise elimination is also not feasible. From last three decades researcher trying to develop new techniques to make better the performance of speech enhancement methods held at different noisy environments. In this paper there is an overview of speech enhancement techniques by using Spectral Subtraction Method (SSM) and Wavelet Packets Method (WPM) .is used as to minimize the noise present in the speech signal [1].

In figure 1 we can see the block diagram of proposed organization. First of all clean speech signals to be taken which is in .wav format. Then add the different noises into this signal to get the noisy speech signal. There are six different noises can be added in the speech signal. Then to enhance this signal remove this noise by using one of the methods like as Spectral Subtraction Method (SSM) and Wavelet Packets Method (WPM). Then finally we get the enhanced speech signal. Lastly, the different parameters of the speech signal are measured.

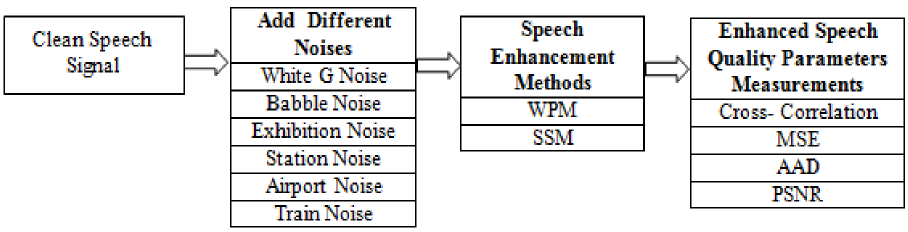

Fig.1. Proposed Block Diagram

The organization of the paper is as follows. Section 2 describes of proposed method Spectral Subtraction Method (SSM). Section 3 describes the proposed Wavelet Packet Method (WPM). Section 4 evaluates the

Measures of Performance Parameters of proposed system. Section 5 shows the results of the proposed system, and final section 6 shows the conclusion of this paper.

II. SPECTRAL SUBTRACTION METHOD (SSM)

The Spectral Subtraction method is one of the methods which are used over a large area for speech enhancement because it has a lower computational load. The input given to this method is noisy speech.

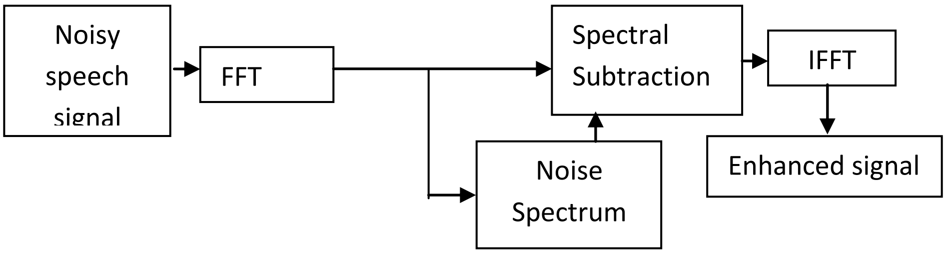

Fig.2. Spectral Subtraction

The windowing technique has used for taking the unchanging number of specimen of the speech signal method that are without a break in nature. The Fourier transform has been applied for converting the signal from frequency domain to time domain, In this paper, fast Fourier transform and a Hamming window is used for windowing and Fourier transform.

For estimating the noise the Minimum Mean Square Error (MMSE) estimator has been used. An estimated noise spectrum has been deducted from the noisy speech input for obtaining clean speech. A magnitude spectrum of a speech signal can be easily restored with the help of this technique [2]. Figure 2 shows the spectral subtraction technique. For the implementation of the MMSE estimator, the probability distribution of the noise Fourier and Speech Fourier expansion coefficients should be known, this is calculated by using MMSE filter**.** To obtain the signal in its time domain form Inverse Fourier transform of the enhanced speech is taken. At this stage the phase of the signal, in its original form, is added to the magnitude. In this way, an enhanced version of the noisy Speech signal is obtained as output.

III. THE WAVELET PACKET METHOD (WPM)

The wavelets theory used in a various fields like, statistics, physics, medicine and biology, etc. In this paper wavelet packet, Coiflet 5 is used for speech enhancement. The Wavelet Packet Decomposition (WPD is a wavelet transform in which the discrete-time signal passes through more filters in compare with a discrete wavelet transform (DWT) [3]. A Wavelet packet de-noising carried in four steps, as given below

A) Wavelet packet decomposition:

In decomposition process, signal S has been decomposed up to N levels. If the human voice is considered and if the more-frequency components are takeoff from the speech signal, then the voice sounds are different**.** Whereas the words present in it are clearly recognized and still audible by the person. But, if enough less frequency components are takeoff, then a resulting audio sign is not audible. Figure 1 determines the stage for signal decomposition. In this, the signal is decomposed over three levels to the high pass filter and low pass filter for calculating a diagonal and approximation components from the signal. Where g[n] is coefficients of low-pass approximation and the h[n] are the high-pass detail coefficients [4]. The Multiple-level decomposition process has been calculated by using the successive approximations. In this one signal is broken down into a many smaller resolution components which are called a wavelet decomposition tree which is shown in figure 2.

B) Entropy:

It is used to calculate optimal trees i.e. based on nature of the signal a particular number of stages are selected. The entropy is one measure of knowledge present in speech signal regularity. There are various types of entropy like as Shannon entropy, P-order standard entropy, log energy entropy, entropy thresholding, and SURE entropy. In this paperwork, the Shannon entropy is applied for detecting the best tree.

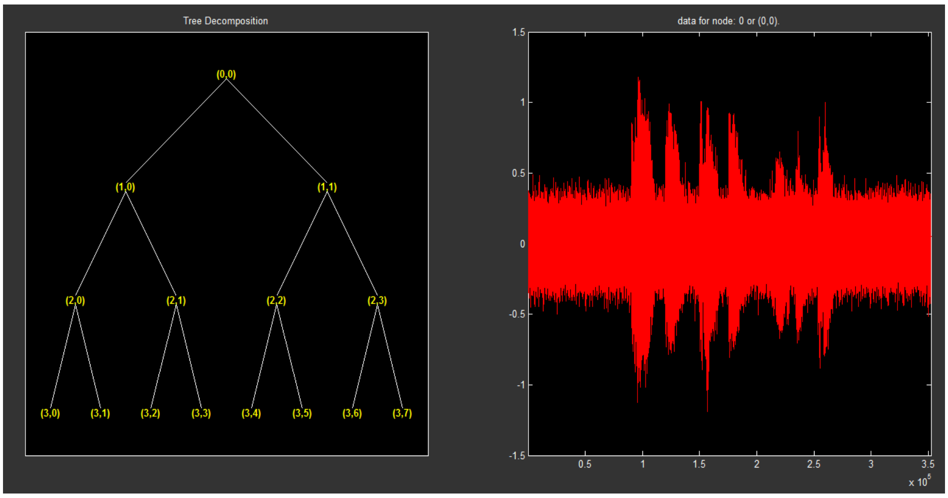

Fig.3. Wavelet Decomposition a)Tree b)Signal

C) Wavelet thresholding:

It is used for the calculation of the threshold values presents in a speech- signal. There are two types of thresholding are found namely: hard and soft. In the hard thresholding a data less than or equal to the threshold value becomes zero, otherwise a data unchanged. In this paper, soft thresholding is used. In the soft thresholding, the absolute value of the signal is compared with the threshold value of the signal. When the absolute value data is less than or equal to the threshold value, it will reduce to zero. If the data it is greater than the threshold value, the data becomes the difference between the data and the threshold value [5]. In equations, 1 and 2 the functions for hard threshold and the soft threshold are shown.

The hard threshold function is defined as

$$w_{j,k} = w_{j,k} \;\ldots\ldots\ldots\ldots\ldots\ldots \quad |w_{j,k}| \geq \lambda$$
$$= 0 \;\ldots\ldots\ldots\ldots \quad |w_{j,k}| < \lambda \qquad (1)$$

The soft threshold function can be expressed as

$$w_{j,k} = [\mathrm{sgn}(w_{j,k})](w_{j,k} - \lambda)\ldots\ldots\ldots\ldots\ldots \; |w_{j,k}| > \lambda$$
$$= 0 \;\ldots\ldots\ldots\ldots\ldots\ldots \; |w_{j,k}| < \lambda \qquad (2)$$

Where, wj,k is the wavelet packet coefficients, wj,k is the estimated wavelet packet coefficients, and λ is a threshold.

D) Wavelet reconstruction

The next step is to reconstruct the wavelet packet. In this method, the N^{th}-high frequency coefficients and N^{th-} low-frequency wavelet coefficients are calculated to reconstruct the wavelet packets. Out of these four steps, the most difficult step is selecting a threshold value and how will quantify this threshold value.

IV. MEASURES OF PERFORMANCE PARAMETERS

This section shows the different parameters measured from the enhanced speech to measure the quality of enhanced speech w.r.t. noisy speech signal.

A) Mean Square Error (MSE)

The MSE used to shows the true values of the quantity being estimated from the signal and determines the difference between the values implied by an estimator. The Mean Square Error shows the wanted value of a squared error loss [6]. This method help to measures the average of squares of the errors and also help to difference between the values implied by estimator and the amount of quantity that is to be estimated gives us an error value of a signal. A Mean Square Error can be represented by using:

$$MSE = \frac{\sum (Enhanced - Input)^2}{}$$

$$Length\ (Input) \qquad (3)$$

B) Average Absolute Distortion (AAD)

It is the AAD which is the warping of the original shape from the signal. The Distortion is always not wanted in a signal and often try to make to reduce it as much as possible [7]. In the paper, we use the parameter Average Absolute Distortion to measure the distortion in add to the speech- signal in comparison with the original signal.

$$AAD = \frac{\sum (Enhanced - Input)}{Length\ (Input)} \qquad (4)$$

C) Cross- Correlation

The co-relation between the two signals, original signal and filtered signal one is the parameter to liken them. In a paper used the command 'crosscorr' from MATLAB, which plots and computes the sample cross-correlation. It is used to find the relation between the two signals.

D) Peak Signal to Noise Ratio (PSNR)

In case of an image compression, for an example, a signal is an original image and the noise is the error between them introduced due to the compression. PSNR is nearly correct value for the human perception of the good quality of reconstruction of speech signal [8]. A higher PSNR shows that the re-constructed signal has the better quality, but the reverse may be true in case of some applications. PSNR can be calculated by:

$$PSNR = 10 \times \log_{10} \frac{Length \times max[Input^2]}{Input^2 - Enhanced^2} \qquad (5)$$

V. RESULTS

This is the section which reports experimental results of considered speech adjustment using transform domain techniques. The techniques as mentioned above are useful to the output waveforms and noisy speech, were evaluated as indicted below.

The figure 4a) determine the original- speech- signal that is also in .wav format, the figure 4b) determine the noisy- speech- signal in which the white -Gaussian noise added.

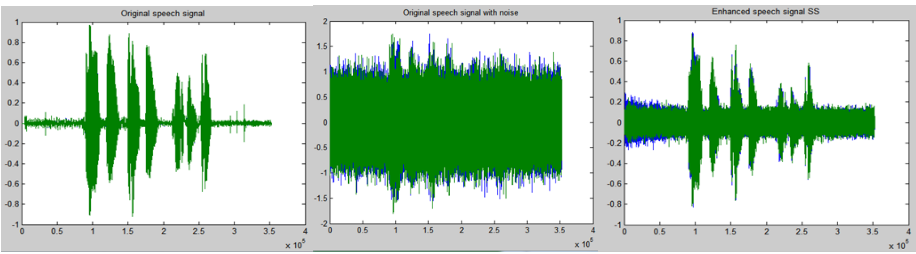

Fig.4. a) Original Speech Signal, b) Original Speech Signal with Noise, c) Enhanced Speech Signal using Spectral Subtraction.

Figure 5 determine the experimental results of the speech- signal. A first graph determine the original- speech -signal that found in .wav format; the second graph determine the noisy speech- signal in which the white -Gaussian noise added, and a third graph shows the enhanced speech signal.

IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 113-118
www.iosrjournals.org

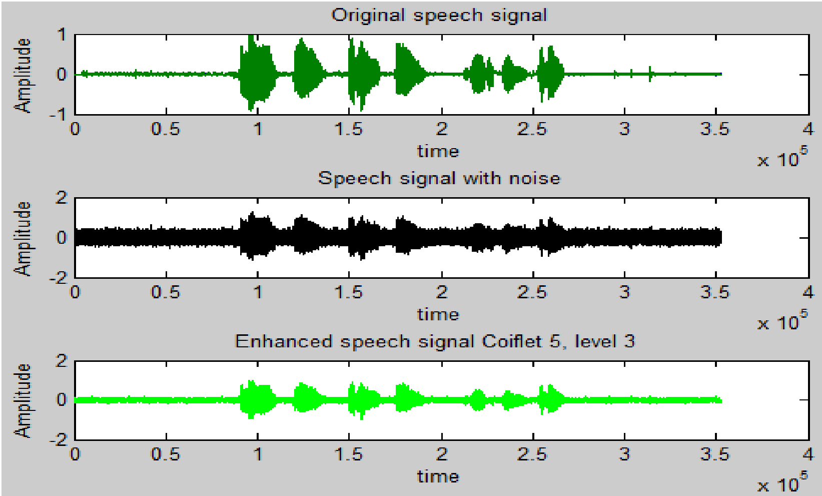

Fig.5. Results of wavelet packet method.

Table number 1 shows experimental results of enhanced speech- signal using SSM and WPM method.

TABLE 1. Results of enhanced speech signal quality measurement

Results for Cross correlation			**Results for AAD**		**Results for MSE (db)**		**Results for PSNR (db)**	
Input-SNR	SSM	WPM	SSM	WPM	SSM	WPM	SSM	WPM
0db	*.2703*	*.9375*	*.00044*	*.00002*	*.0298*	*.0019*	*14.95*	*27.00*
2db	*.3765*	*.9572*	*.00024*	*.00012*	*.0192*	*.001200*	*16.87*	*28.79*
5db	*.5604*	*.9753*	*.00034*	*.00013*	*.0113*	*.000693*	*19.16*	*31.28*
10db	*.8081*	*.9885*	*.00029*	*.00012*	*.0048*	*.000318*	*22.89*	*34.67*
15db	*.9267*	*.9927*	*.00003*	*.00001*	*.002*	*.000201*	*26.66*	*36.67*
20db	*.9694*	*.994*	*.00001*	*.00001*	*.00093*	*.000164*	*30.03*	*37.54*

Figure 6 a) shows the Comparison of cross correlation for spectral subtraction and wavelet packets. Figure 6 b) shows the Comparison of AAD for spectral subtraction and wavelet packets. Figure 7 a) shows the Comparison of MSE for spectral subtraction and wavelet packets. Figure 6 b) shows the Comparison of PSNR for spectral subtraction and wavelet packets.

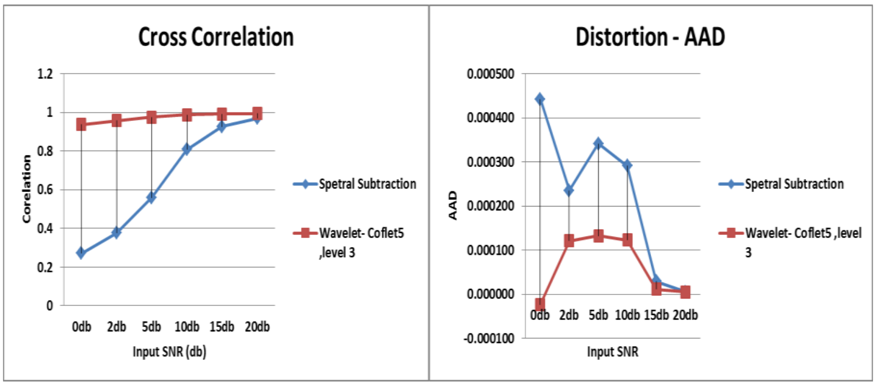

Fig.6 a). Comparison of cross correlation for spectral subtraction and wavelet packets, b)Comparison of AAD for spectral subtraction and wavelet packets.

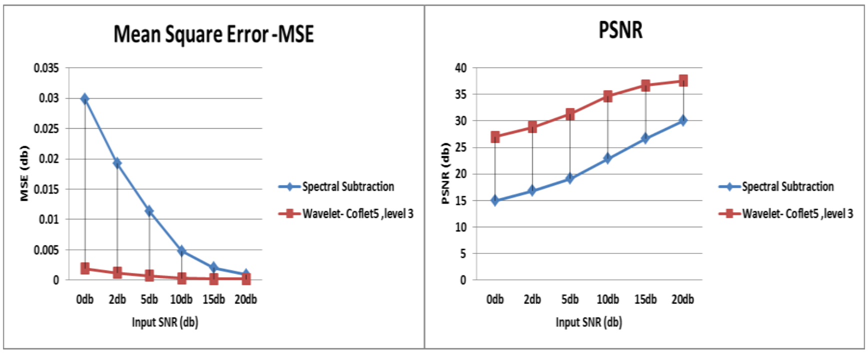

Fig.7 a). Comparison of cross MSE wavelet packets, b) Comparison of PSNR for spectral subtraction and wavelet packets.

VI. CONCLUSION

We provide a practical approach in how to put into practice the wavelet packets in noisy speech data to improve clarity and signal retrieval. From the graphs, we can clearly see and eventually conclude that Wavelet packets method provides the better results compared to spectral subtraction method. It also filters out the noise and renders us with clean speech in a more flat line Peak SNR value, less distortion in a varying loudness of the speech.

REFERENCES

[1]. Islam, M.T.; Shahnaz, C.; Fattah, S.A., "Speech enhancement based on a modified spectral subtraction method," Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on , vol., no., pp.1085,1088, 3-6 Aug. 2014 doi: 10.1109/MWSCAS.2014.6908607.

[2]. Gupta, V.K.; Bhowmick, A.; Chandra, M.; Sharan, S.N., "Speech Enhancement Using MMSE Estimation and Spectral Subtraction Methods," Devices and Communications (ICDeCom), 2011 International Conference on , vol., no., pp.1,5, 24-25 Feb. 2011,doi: 10.1109/ICDECOM.2011.5738532.

[3]. Goodarzi, H.M.; Seyedtabaii, S., "Speech Enhancement Using Spectral Subtraction Based on a Modified Noise Minimum Statistics Estimation," INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference on , vol., no., pp.1339,1343, 25-27 Aug. 2009,doi: 10.1109/NCM.2009.272.

[4]. Bing-yin Xia; Yan Liang; Chang-chun Bao, "A modified spectral subtraction method for speech enhancement based on masking property of human auditory system," Wireless Communications & Signal Processing, 2009. WCSP 2009. International Conference on , vol., no., pp.1,5, 13-15 Nov. 2009. doi: 10.1109/WCSP.2009.5371466.

[5]. Udrea, R.M.; Ciochina, S., "Speech enhancement using spectral over-subtraction and residual noise reduction," Signals, Circuits and Systems, 2003. SCS 2003. International Symposium on , vol.1, no., pp.165,168 vol.1, 0-0 2003,doi: 10.1109/SCS.2003.1226974.

[6]. Hasan, M.K.; Salahuddin, S.; Khan, M.R., "A modified a priori SNR for speech enhancement using spectral subtraction rules," Signal Processing Letters, IEEE , vol.11, no.4, pp.450,453, April 2004,doi: 10.1109/LSP.2004.824017.

[7]. Guangyan Wang; Xia Wang; Xiaoqun Zhao, "Speech enhancement based on a combined spectral subtraction with spectral estimation in various noise environment," Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on , vol., no., pp.1424,1429, 7-9 July 2008,doi: 10.1109/ICALIP.2008.4590225.

[8]. Lan Xu; Hon Keung Kwan, "Adaptive wavelet denoising system for speech enhancement," Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on , vol., no., pp.3210,3213, 18-21 May 2008,doi: 10.1109/ISCAS.2008.4542141.