

A New Robust Hybrid Approach to Enhance Speech in Mobile Communication Systems

¹Manimegalai Govindan Sumithra, ²Keppana Gounder Thanuskodi and ³Bharathi Deepa

¹Department of Electronics and Communication Engineering,
Bannari Amman Inst. of Technology,

²Department of Electrical and Electronics,
Engineering Akshaya College of Engineering and Technology,

³Department of Electronics and Communication Engineering,
Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India

Abstract: Problem statement: The received voice signal in mobile communication is often disturbed by background noise and hence there is a need for good noise reduction methods for enhancing speech. It is well known that denoising is a compromise between the removal of the largest possible amount of noise and the preservation of signal integrity. To address this issue, a new method for enhancing speech from background interference is presented in this study by fusing dual band spectral subtraction with adaptive noise estimator and wavelet packet based thresholding method. **Approach:** The proposed system uses the combination of dual band Spectral Subtraction method with adaptive noise estimator for pre-processing, in order to initially reduce the noise level and further the quality of speech is improved by Wavelet Packet Transform (WPT) based level dependent thresholding method. The threshold value is determined by using Stein's Unbiased Risk Estimator (SURE) and hard, soft, Garrotte, μ -law and a proposed modified soft thresholding functions are considered for denoising. **Results:** The proposed method was investigated by ten different clean speech samples (five male and five female) taken from TIMIT database and thirteen different noise sources to degrade the speech artificially and the energy level of the noise is scaled such that the overall SNR of the noisy speech is maintained at -5, 0, 5, 10 and 15 dB and finally the results are evaluated using objective and subjective measures. **Conclusion/Recommendations:** It is suggested from the experimental results that the proposed scheme gives improved spectral performance, reflects in better speech quality in all types of noisy environment. For better speech enhancement in noise dominated regions, the system efficiency is further improved by fusing threshold values for wavelet denoising.

Keywords: Speech enhancement, dual band spectral subtraction, Wavelet Packet Decomposition (WPT), stein's unbiased risk estimate (SURE)

INTRODUCTION

In Communication Systems, speech signals can be contaminated by environmental noise and, as a result, the communication quality can be affected making the speech less intelligible. The fast growing mobile communication of today demands increasingly better sound quality of the received speech signal. Disturbances that make the speech less intelligible often come from the background environment such as a car engine or humming people. Voice quality and intelligibility are always important for communication systems, either wired or wireless. Speech enhancement algorithms have therefore attracted a great deal of

interest in the past two decades. It is well known that denoising is a compromise between the removal of the largest possible amount of noise and the preservation of signal integrity (Ramadan, 2008). In mobile communication systems, the performance of the speech coder weakened by undesirable background noises is an annoying problem. One way to reduce this problem is to apply a speech enhancement step to improve the system performance of voice communication in the presence of ambient noises (Michael *et al.*, 2007). This is important in a variety of contexts, such as in environments with interfering background noise and in speech recognition systems, hands free environment for cars, hearing aids. The effectiveness of the speech

Corresponding Author: Manimegalai Govindan Sumithra, Department of Electronics and Communication Engineering, Bannari Amman Inst. of Technology, Sathyamangalam, Tamil Nadu, India

enhancement system can be measured based on how well it performs in light of the trade-off, maintained between distortions in the processed speech and the amount of noise suppressed.

Literature review: Existing approaches to this task include traditional methods such as wiener filtering, spectral subtraction and Ephraim Malah filtering. When the noise or the signal-to-noise ratio (SNR) is known, contemporary techniques can yield quasi-optimal solutions to the problem of denoising. Spectral subtraction is a non-parametric method which requires only an estimate of the noise spectrum, in order to obtain the original clean speech. Since the noise spectrum is estimated from the pause periods and used for the whole data, spectral subtraction is suitable for stationary noises or very slowly varying noises so that the change in the noise power spectrum can be updated. Modified Spectral Subtraction (MSS) method is introduced, to prevent destructive subtraction of the speech during the removal of residual noise. It was based on identifying and enhancing speech regions in the noisy speech signal. Further to reduce the distortions in the speech signal Multiband Spectral Subtraction (MBSS) is introduced for maintaining a high level of speech quality. The added computational complexity of the algorithm is minimal. Also results shows that four linearly-spaced frequency bands were adequate in obtaining good speech quality. A drawback of these enhancement techniques is the necessity to estimate the noise or the SNR.

Recently, the effective noise suppression is achieved by transforming the noisy signal into the wavelet domain and preserving only the local maxima of the transform (Michael *et al.*, 2007; Johnson *et al.*, 2007; El-Leithy and Sheta, 2009). Wavelet denoising is commonly used for speech enhancement because of the simplicity of its implementation. The effectiveness of wavelet-based de-noising is due to the fact that, for a wide variety of signal classes, the energy of the signal gets packed in few relatively large coefficients, while the noise energy is spread over a larger number of coefficients. One of the main advantages of wavelet denoising is that it does not require any assumptions about the noisy signal and can deal with signals with discontinuities and spatial variations. An improved wavelet-based speech enhancement method using the perceptual wavelet packet decomposition and the Teager energy operator was more suitable for real noise cases (Chen and Wang, 2004). The main advantage of this is that the over thresholding of speech segments can be avoided. As a consequence, the enhanced speech quality can be increased substantially from those of

conventional approaches. In addition, it does not require complete estimation of noise level of the SNR. A new speech enhancement system using the wavelet thresholding algorithm is presented Sumithra *et al.* (2009). A novel algorithm of wavelet coefficient threshold (WCT) based on time-frequency adaptation is introduced in. In addition, an unvoiced speech enhancement algorithm is also integrated into the system to improve the intelligibility of speech. The wavelet coefficient threshold (WCT) of each sub band is first temporally adjusted according to the value of a posterior signal-to-noise ratio (SNR). To prevent the degradation of unvoiced sounds during noise, the algorithm utilizes a simple speech/noise detector (SND) and further divides speech signal into unvoiced and voiced sounds (Wang, 2010). Then, appropriate wavelet thresholding is applied according to voiced/unvoiced (V/U) decision. Based on the masking properties of human auditory system, a perceptual gain factor is adopted into wavelet thresholding for suppressing musical residual noise.

The main objective of the proposed method is to improve on existing single-microphone schemes for an extended range of noise types and noise levels, thereby making this method more suitable for mobile speech communication applications than the existing

The study presented here uses the combination of spectral subtraction and wavelet packet based threshold method for speech enhancement with the idea that the improved representational capability of the proposed method on speech signals could lead to better separation of signal and noise components within the coefficients and therefore better enhancement results and to improve on the existing single-microphone schemes for an extended range of noise types and noise levels in real environment, thereby making this method more suitable for mobile speech communication applications than the existing. The proposed scheme consists of two parts. First part of this scheme performs pre-estimation of speech using dual band Spectral Subtraction. Second part introduces a speech enhancement system based on a Wavelet Packet Transform. The performance of the proposed method was evaluated on several speakers and under various adverse noise conditions. The obtained results of the proposed method shows that it is well suited for real world noise conditions and yields better spectral performance.

MATERIALS AND METHODS

The proposed SSWPT system structure is shown in Fig. 1. In order to initially reduce the noise level, the noisy speech is first pre processed with a dual band

spectral subtraction routine with adaptive noise estimator. A three level wavelet packet transform is then applied to decompose the noisy signal into sub bands. To account for non-stationary and correlated noise, thresholds are independently estimated for each time frame and wavelet decomposition sub band. This is further refined using a modified soft thresholding approach based on a SURE risk rule. Finally, the inverse wavelet packet transform synthesizes the enhanced speech.

Pre-processing: Pre-processing is done using dual spectral subtraction and it consists of four stages. In the first stage, the signal is windowed and the magnitude spectrum is estimated using the FFT. In the second stage, split the noise and speech spectra into different frequency bands and calculate the over-subtraction factor for each band.

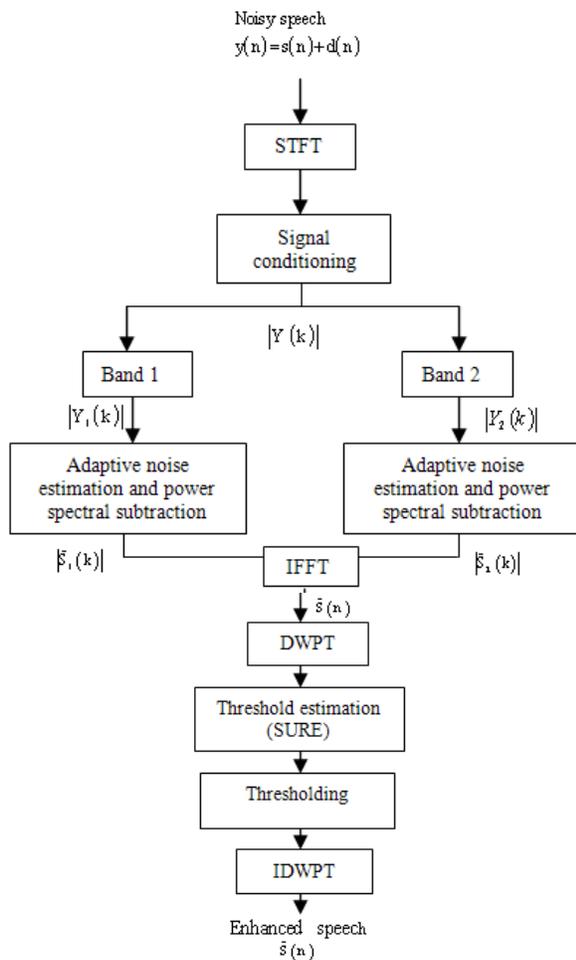


Fig. 1: Block diagram for proposed method

The third stage includes processing the individual frequency bands by subtracting the corresponding noise spectrum from the noisy speech spectrum. Lastly, the modified frequency bands are recombined and the time signal is obtained by using the noisy phase information and taking the IFFT in the fourth stage. The effect of signal conditioning operations is to neutralize the distortion in the spectral content of the input data due to the analysis window and to precondition the input data to surmount the distortion due to errors in the subtraction process. To reduce the effect of residual noise in the enhanced speech, it is necessary to reduce the variance of the frequency content of the signal. Hence, instead of directly using the power spectra of the signal, a smoothed version of the power spectra can be used. However, it is seen that smoothing (Local or magnitude averaging) of the estimated noise spectrum is helpful in reducing residual noise. Assuming the additive noise to be stationary and uncorrelated with the clean speech signal, the resulting input corrupted speech can be expressed as:

$$y(n) = s(n) + d(n) \tag{1}$$

The estimate of the clean speech spectrum in the i^{th} band is obtained by:

$$|\hat{S}_i(k)|^2 = |\bar{Y}_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2 \tag{2}$$

$b_i \leq k \leq e_i, i = 1, 2$

Where b_i and e_i are the beginning and ending frequency bins of the i^{th} frequency band, δ_i is an additional band-subtraction factor that can be individually set for each frequency band to customize the noise removal process and α_i is band specific over subtraction factor.

Adaptive noise estimation: Noise estimation plays an important role in this method of speech enhancement (Rangachari and Loizou, 2006). The algorithm used for noise estimation in this study is based on updating the noise estimate by tracking the silence regions of speech. But, the noise estimate is updated continuously in every frame irrespective of speech present or absent frames. This is based on the concept that the power spectrum of speech was both localized in time and frequency, i.e. even in the speech present frames only a fraction of the entire frequency spectrum. The noise spectrum estimate is updated using the following recursive equation:

$$D(\alpha, k) = \delta_s(\alpha, k) D(\alpha - 1, k) + (1 - \delta_s(\alpha, k)) |Y(\alpha, k)|^2 \tag{3}$$

where

- $D(a,k)$ = Estimate of the noise power spectrum
- $\delta(a,k)$ = Frequency dependent smoothing factor

The value of δ is taken to be equal to 1 for speech present frame during high speech activity and is set to be 0.8 for speech present frame.

Wavelet packet decomposition: The conventional wavelet transform decomposes only the low frequency components to obtain the next level's approximation and detail components; the current level of the detail components remains intact (Ghanbari and Karami-Mollaei, 2006) Thus the computation of Discrete Wavelet Transform (DWT) is providing sufficient information for both analysis and synthesis of the original signal, with a significant reduction in the computation time (Wang, 2010). The DWT is considerably easier to implement without needing to perform numerical integration as Continuous Wavelet Transform (CWT).

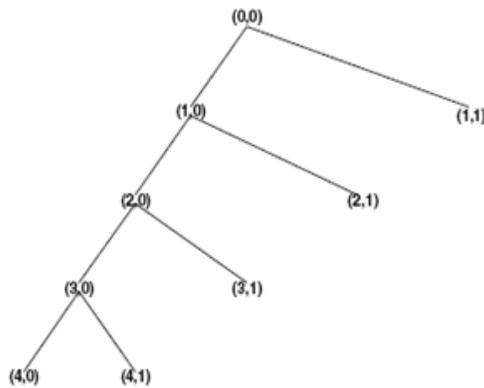


Fig. 2: DWT Computation

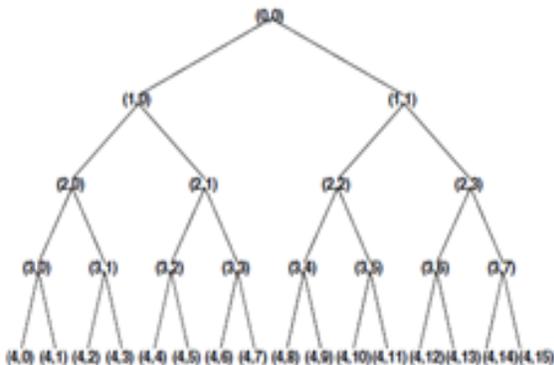


Fig. 3: DWPT Computation

DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high pass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive high pass and low pass filtering of the time domain signal. In the DWT, each level is calculated by passing the previous approximation coefficients through high and low pass filters. For n levels of decomposition DWT produces $(n+1)$ sets of coefficients. Figure 2 represents the filter bank decompositions, with the left and right branches at each node representing a matched pair of low-pass and high-pass wavelet filters followed by down sampling.

This results in more components representing the signal and provides more flexibility, which makes the improvement in noise reduction and spectral performance. The depth of the Wavelet Packet Tree shown in Fig. 3, can be varied over the available frequency range, resulting in configurable filter bank decomposition. The decomposition of both approximations and details generates a wavelet packet. This results in a balanced binary tree structure. The root of the tree is the original dataset The next level of the tree is the result of one step of wavelet transform. Subsequent levels in the tree are constructed by recursively applying the wavelet transform step to the low and high pass filter results from the previous wavelet transform step.

However, in the wavelet packet analysis (Ayache *et al.*, 2010), both the approximation and details at a certain level are further decomposed into the next level, which means the wavelet packet analysis can provide a more precise frequency resolution than the wavelet analysis. This idea has been used to create customized Wavelet Packet Transforms where the filter banks match a perceptual auditory scale, such as the Bark scale, for use in speech representation, coding and enhancement (Boubakir and Berkani, 2010). The use of bark-scale WPT for enhancement has so far indicated a small but significant gain in overall enhancement quality due to this perceptual specialization. This perceptual WPT, using auditory critical band scaling following as shown in Fig. 4, is implemented in this study as a reference method for comparison to the new technique. Similarly the inverse wavelet packet can reconstruct the original signal from the wavelet packet decomposed spectrum. The inverse wavelet packet is done starting from the coarsest decomposition level where the WPT coefficients are up sampled before passing through a pair of reconstruction filters. Note that, the wavelet that is used as a base for decomposition cannot be changed if we want to reconstruct the original signal. Dabuchies 14 tap wavelet has been chosen for used for denoising.

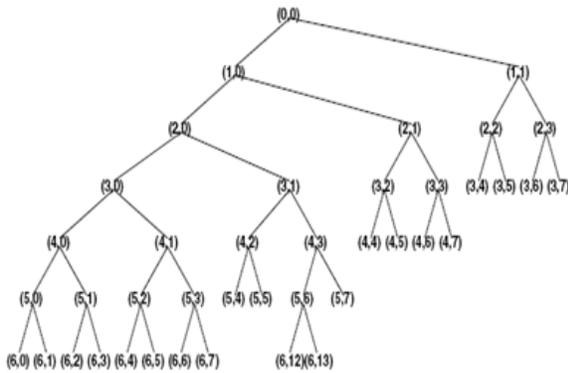


Fig. 4: PWPT Computation

Wavelet packet denoising: For the applications of interest, noise is primarily high frequency, while the signal of interest is primarily low frequency. Because the wavelet transform decomposes the signal neatly into approximation (low frequency) and detail (high frequency) coefficients, the detail coefficients will contain much of the noise (Mahesh *et al.*, 2010; O’Shaughnessy, 2005). This suggests a method for denoising the signal: simply reduce the size of the detail coefficients before using them to reconstruct the signal. This approach is called thresholding or shrinkage the detail coefficients. Of course, the detail coefficients entirely cannot be thrown away, they still contain some important features of the original signal. A generalization of the discrete wavelet transform is the discrete wavelet packet transforms (DWPT) which keeps splitting both low pass and high pass sub-bands at all scales in the filter bank implementation, thus Wavelet Packet obtains a flexible and a detail analysis transform. So the Wavelet Packet transform is used for de-noising. The main steps of signal denoising are : (1).Wavelet packet transform of pre estimated speech signal. (2).Shrinkage of the empirical wavelet coefficients. (3). Inverse wavelet packet transform of the modified coefficients.

The denoising procedure requires the estimation of the noise level. In this study Stein's Unbiased Estimate of Risk (SURE) (Mahesh *et al.*, 2010; Hu and Loizou, 2004) has been chosen as a principle for selecting a threshold to be used for denoising. SURE is an adaptive threshold selection rule. It is data driven. The aim of estimate is to minimize the risk. Because the coefficients of true signal are unknown, the true risk is also not unknown. This technique calls for setting the level dependent threshold T to where $N_{j,k}$ is the number of the samples in the node (j,k) scale j and $C_{j,k}$ represents high frequency wavelet coefficients which are

used to identify the noise components at j^{th} level decomposition and sub-band k in the wavelet packet tree.

Selection of threshold function: Obviously, the choice of threshold directly influences the effectiveness of the denoising algorithm. Too high a threshold would result in too many wavelet packet decomposition coefficients being reset as zero and thus destroying too many details of the signal, while with too low a threshold the expected denoising effect could not be achieved. Various kinds of thresholding have been proposed in literature and which kind of thresholding is best depends on the application. The two different approaches which are usually applied to denoise the signals are hard thresholding and soft thresholding (Hu and Loizou, 2007).The soft thresholded signal can be written as:

$$Thr_{soft}(X, T) = \begin{cases} \text{sgn}(X) (|X| - T) & |X| > T \\ 0 & |X| \leq T \end{cases} \quad (5)$$

$$Thr_{hard}(X, T) = \begin{cases} X & |X| > T \\ 0 & |X| \leq T \end{cases} \quad (6)$$

where X represents the wavelet coefficients before thresholding and T is the threshold. According to Donoho (Jiang *et al.*, 2006) the wavelet soft thresholding method achieves asymptotically near-optimal minimax MSE over a wide range set of functions with certain smoothness. The hard thresholding function zeroes out all coefficients with magnitude smaller than the threshold value. The hard thresholding method is reported to have better MSE than soft thresholding in some situations where the signals to be de-noised have a significant number of large detail coefficients. Soft and hard thresholding methods suffer from distortion of the speech because they set coefficients to zero that may carry useful information, resulting in observable sharp time frequency discontinuities in the speech spectrogram. In addition to the above thresholding functions, μ -law, Garrote and modified soft thresholding (Ali *et al.*, 2010) functions are also considered for analysis to threshold wavelet packet coefficients. The mathematical representation of above functions are shown below:

$$Thr_{Garrote}(X, T) = \begin{cases} 0 & |X| \leq T \\ X - \frac{T^2}{X} & |X| > T \end{cases} \quad (7)$$

$$\text{Thr}_{\text{modifiedsoft}}(T_1, T_2) = \begin{cases} 0 & |X| \leq T_1 \\ \text{sgn}(T) \frac{T_2(|X| - T_1)}{T_2 - T_1} & T_1 \leq |X| \leq T_2 \\ X & T_2 \leq |X| \end{cases} \quad (8)$$

$$\text{Thr}_{\mu\text{-law}}(X, T) = \begin{cases} X & |X| > T \\ T \left(\frac{\left[\frac{(1 + \mu)^{|X|/T} - 1}{\mu} \right] \text{sgn}(X)}{\mu} \right) & |X| < T \end{cases} \quad (9)$$

RESULTS AND DISCUSSION

Two types of experiment is conducted one is under AWGN condition and the next one is under real life noise conditions. In the first type of experiment the clean speech utterance is artificially degraded by adding white Gaussian noise at the following SNR levels in dB:-5,0, 5,+10,+15. Secondly, those utterances are

corrupted at the same SNR levels by adding with pink noise, multiple talkers' noise (Babble noise), HF channel noise, train noise, street noise, factory noise, exhibition noise, f-16 cockpit noise, car noise, airport noise, station noise and restaurant noise in order to investigate how those methods deal with non-stationary real-life noises. The noise corrupted sentences are processed by the proposed method with different threshold selection functions like hard, soft, modified soft, Garrote and μ -law .

Figure 5 shows the time domain and spectrogram representation of Clean, Noisy speech (degraded by Factory noise at 0 dB SNR level), Pre-estimated speech, Enhanced speech using SSWPT with Hard, Soft, μ -law, Garrote and modified soft thresholding respectively. From Fig, 4 it is observed that nearly similar spectral performance is obtained in both soft and Garrote thresholding and improved and comparable spectral performance with clean speech is shown by modified soft thresholding. Also here the performance evaluation is done by both subjective and objective measures.

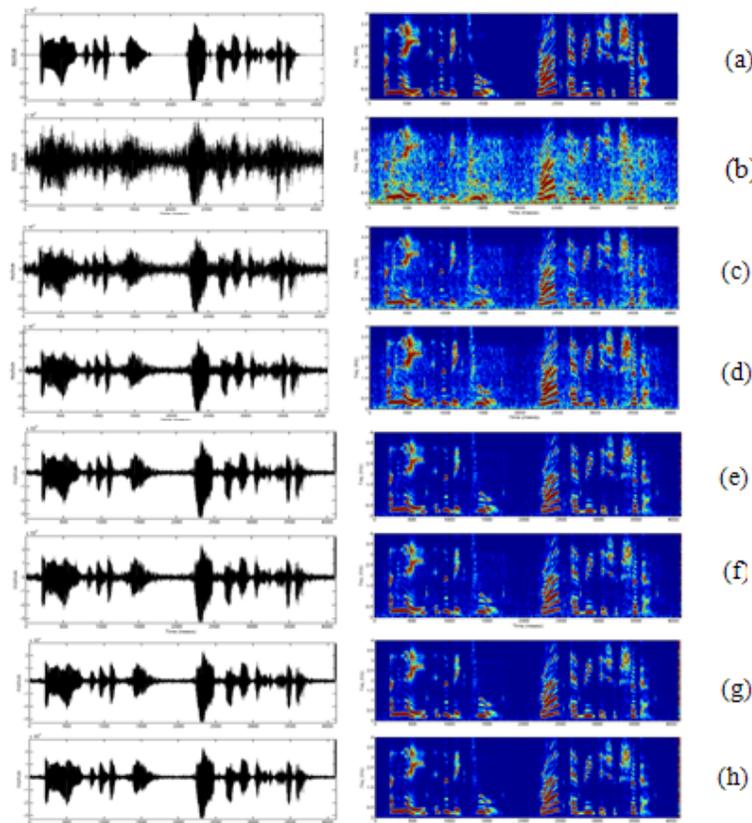


Fig. 5: Time domain and Spectrogram representation of (a)Clean (b) Noisy speech (Train noise 0 dB) (c) Pre-estimated speech , Enhanced speech using SSWPT with (d)Hard (e) Soft (f) μ -law (g) Garrote (h) Modified Soft thresholding

Objective measure: Objective quality measures are based on a mathematical comparison of the original and processed or enhanced speech signals. The two main factors in selecting an objective distortion measure are its performance and complexity (Sumithra *et al.*, 2009; Chomphan, 2010). The parameters considered for evaluating the enhancement algorithms are Signal to Noise Ratio (SNR), Segmental SNR, Minimum Mean Square Error (MMSE) and Spectral distance measure that is Itakuro Saito (IS) distance measure.

Signal to Noise Ratio (SNR): The SNR is a measurement method based on an additive noise model, where the noisy signal $x(n)$ is a superposition of the clean signal $y(n)$ and the additive error $e(n)$ and the global SNR (Supaththara *et al.*, 2010; Stark and Barkana, 2010) is calculated mathematically by:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left[\frac{\sum_n y^2(n)}{\sum_n [y(n) - \hat{y}(n)]^2} \right] \quad (10)$$

Where

$y(n)$ = Clean speech
 $\hat{y}(n)$ = Estimated speech

If the summation is performed over the whole signal length, the operation is called as global SNR.

Minimum Mean Square Error: Mean Square Error (MSE) is defined as the average power of the difference between the enhanced speech and clean one (Chavan *et al.*, 2010; Helmy and Taweel, 2010). It can be obtained by;l:

$$\hat{\sigma} = E \left\{ [y(n) - \hat{y}(n)]^2 \right\} \quad (11)$$

For a better estimation of any signal, MMSE value should be low.

Itakuro-Saito (IS) distance measure: It is a meaningful measure of performance when the two waveforms differ in their phase spectra:

$$d(a, b) = \frac{(a - b)^T R (a - b)}{a^T R (a)} \quad (12)$$

where 'a' is the vector for the prediction coefficients of the clean speech signal, vector R is the (Toeplitz) autocorrelation matrix of the clean speech

signal and vector 'b' is the prediction coefficients of the enhanced signal. Many reported experiments confirmed that two spectra would be perceptually nearly identical if the distance is from 1-10, with lower values indicating lesser distance and better speech quality.

Subjective measure: Subjective Quality measures provide a broad measure of performance since a large difference in quality is necessary to make it distinguishable to the listener. (1)Mean opinion score: To determine MOS, a number of listeners rate the quality of test sentences read aloud over the communications circuit by male and female speakers. A listener gives each sentence a rating as follows: (1) Bad, (2) Poor, (3) Fair, (4) Good, (5) Excellent. The MOS is the arithmetic mean of all the individual scores and can range from 1 (worst) to 5 (best).A program in visual basic is used to collect mean opinion scores from more than ten listeners. All the instructions for the listeners are provided in the program with twelve samples of various db levels.Enhanced speech results across varying realistic noise conditions at different SNR, using baseline and proposed method were analyzed in time domain as well as in spectral domain, shown in Fig. 6 for HF Channel noise. From the Fig. 6 it is evident that the enhanced speech obtained using the proposed and EMF more comparable than others. But from the spectrogram analysis (Helmy and Taweel, 2010) it is obvious that the proposed method yields better result than EMF. But better spectral performance was shown by SSWPT.

Performance comparison under AWGN noise condition: Output SNR results for AWGN noise condition across range of SNR values are shown in Fig. 7. From Fig. 7 it is observed that, for white noise the proposed method with modified thresholding yields higher performance than other thresholding functions even in the case of noise dominated speeches. At 0 dB input SNR level the Garrote thresholding shows higher performance than modified thresholding function. The modified thresholding gave about 9 db improvement at the lower input SNR (-5dB), increasing to about 21 db improvement at the higher input SNR (15dB). Garrote thresholding gave about 7.5 db improvement at the lower input SNR (-5dB), increasing to about 20 db improvement at the higher input SNR(15dB) but it gave 1dB higher than modified thresholding at 0 dB input SNR . Based on these results it cannot be possible to say that particular thresholding is better for denoising. It could be possible after the spectral analysis. Comparison under real life noise conditions.

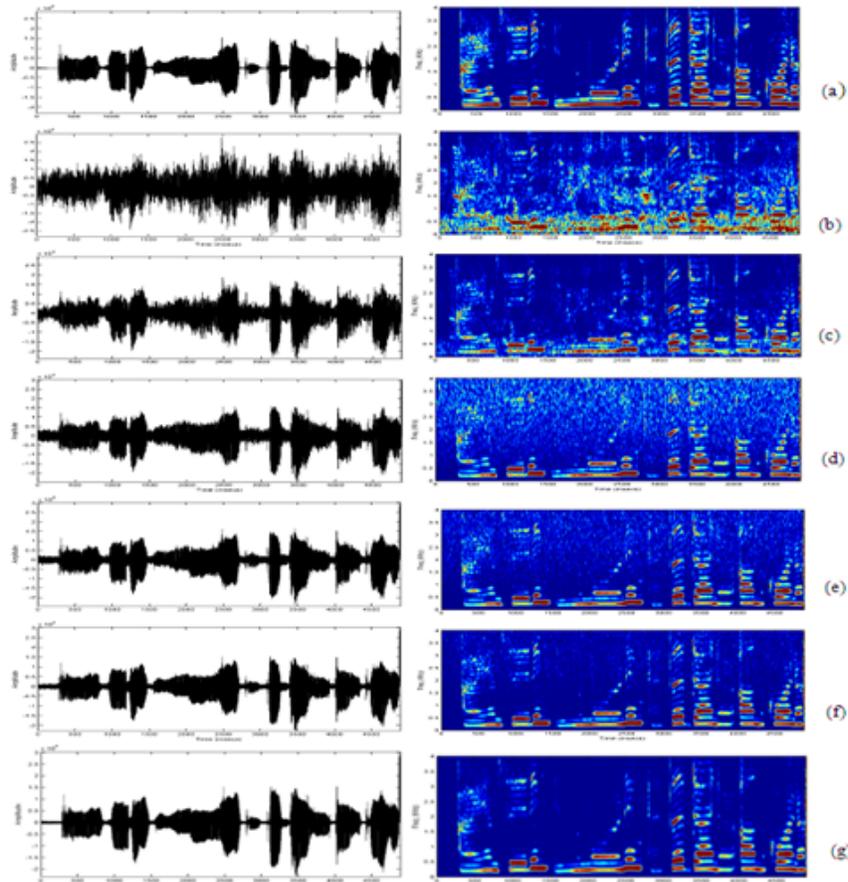


Fig. 6: Time domain and Spectrogram representation of (a) Clean (b) Noisy speech (HF Channel Noise 0 dB) Enhanced speech using (c) SS (d) IWF (e) EMF (f) PWPT (g) SSWPT

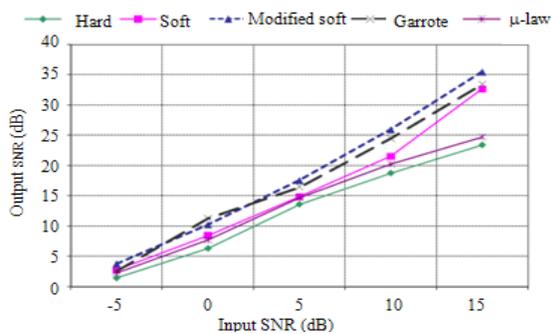


Fig. 7: Comparison of output SNR results under white noise condition

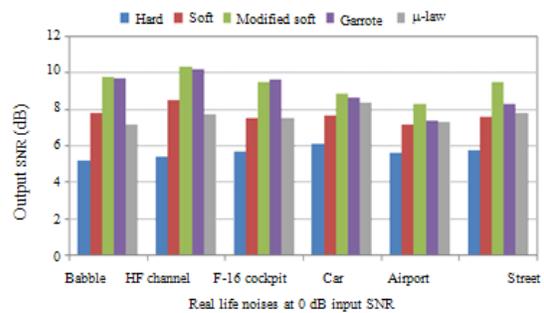


Fig. 8: Comparison of output SNR results under real life noise condition

Figure 8 represents the output SNR results comparison under real life conditions, implies that the results of modified soft and Garrote thresholding are comparable in case of F-16 cockpit noise and in other cases modified soft is showing good performance.

Performance comparison of objective and subjective measure: The performance comparison of objective and subjective measures for clean speech signal corrupted by white Gaussian noise and also other type of noises like pink noise, multiple talkers' noise (Babble noise), HF channel noise, train noise, street

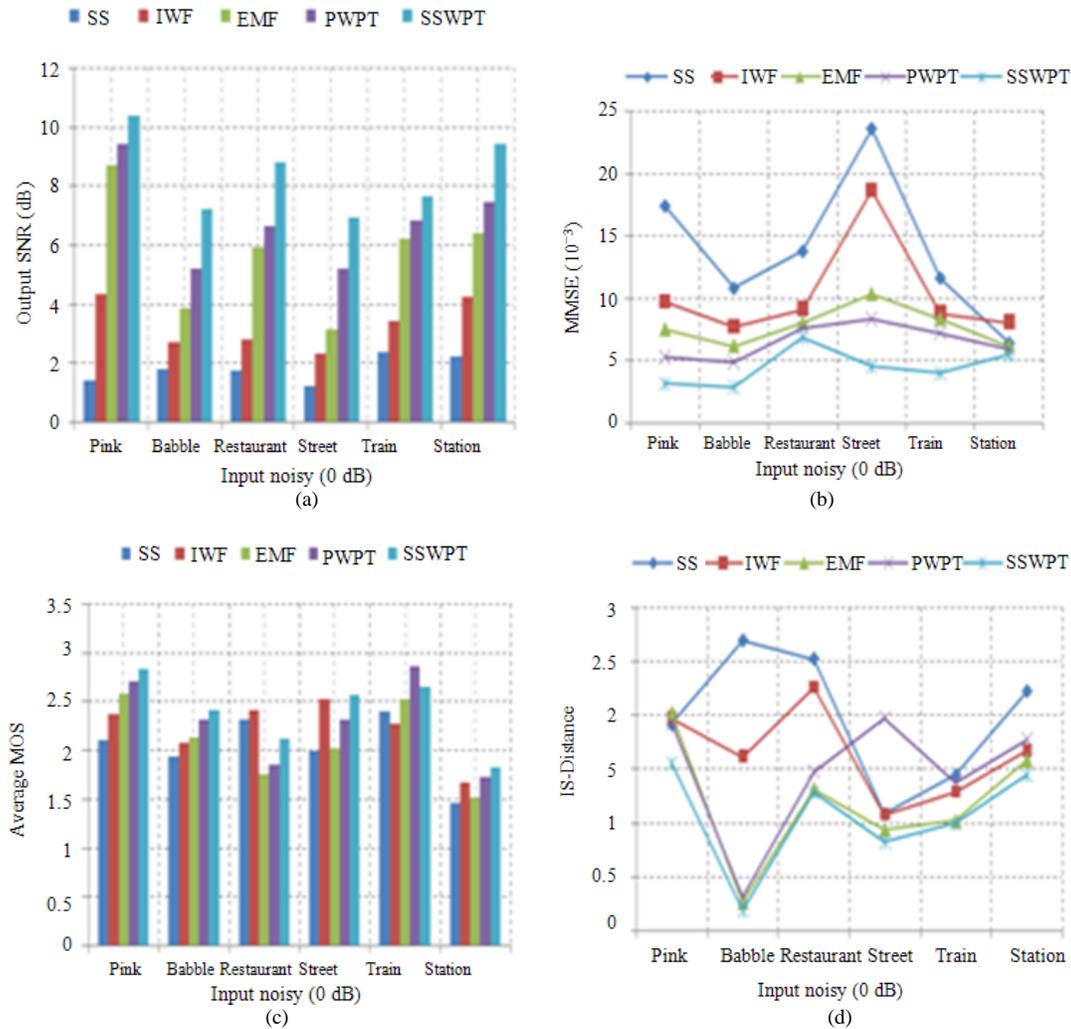


Fig. 9: Comparison of average performance metrics for various noises for different methods at 0dB SNR levels (a) Output SNR (b) average MOS (c) MMSE (d) IS-Distance

noise, factory noise, exhibition noise, f-16 cockpit noise, car noise, airport noise, station noise and restaurant noise can be made at different SNR levels ranging from dB:-5,0,+5,+10,+15. The global SNR improvement across varying realistic noise condition at 0dB SNR is shown in Fig. 9(a). Here the results are given as net improvement, so that relative effectiveness can be seen for all six noise conditions as a function of enhancement method. SSWPT substantially outperforms the other methods in nearly all cases. PWPT approach outperforms the remaining three methods.

Subjective results using Mean Opinion Scores (MOS) for the same noise conditions are shown in Fig. 9(b) where the relative MOS results of PWPT and SSWPT is in line and EMF and IWF scores are

competitive to each other for all presented noise types except train noise where SS yields better score than IWF. Figure 9(c) represents the MMSE value comparison, where SSWPT shows the minimum value and SS gives the maximum value. From the Fig. 9(d) it is observed that the phase difference between the enhanced speech and clean one is having the maximum average value of 2.61 in SS method and minimum average value of 1.06 in SSWPT method.

As can be seen from the Fig. 10(a), it is observed that the average output SNR measure of SSWPT is showing superior performance than PWPT, EMF, IWF and SS for six different types of noises as considered. From the Fig. 10(b) it is inferred that the average MOS score for the proposed method is high when compared to existing. In case of F-16 Cockpit noise the scores of

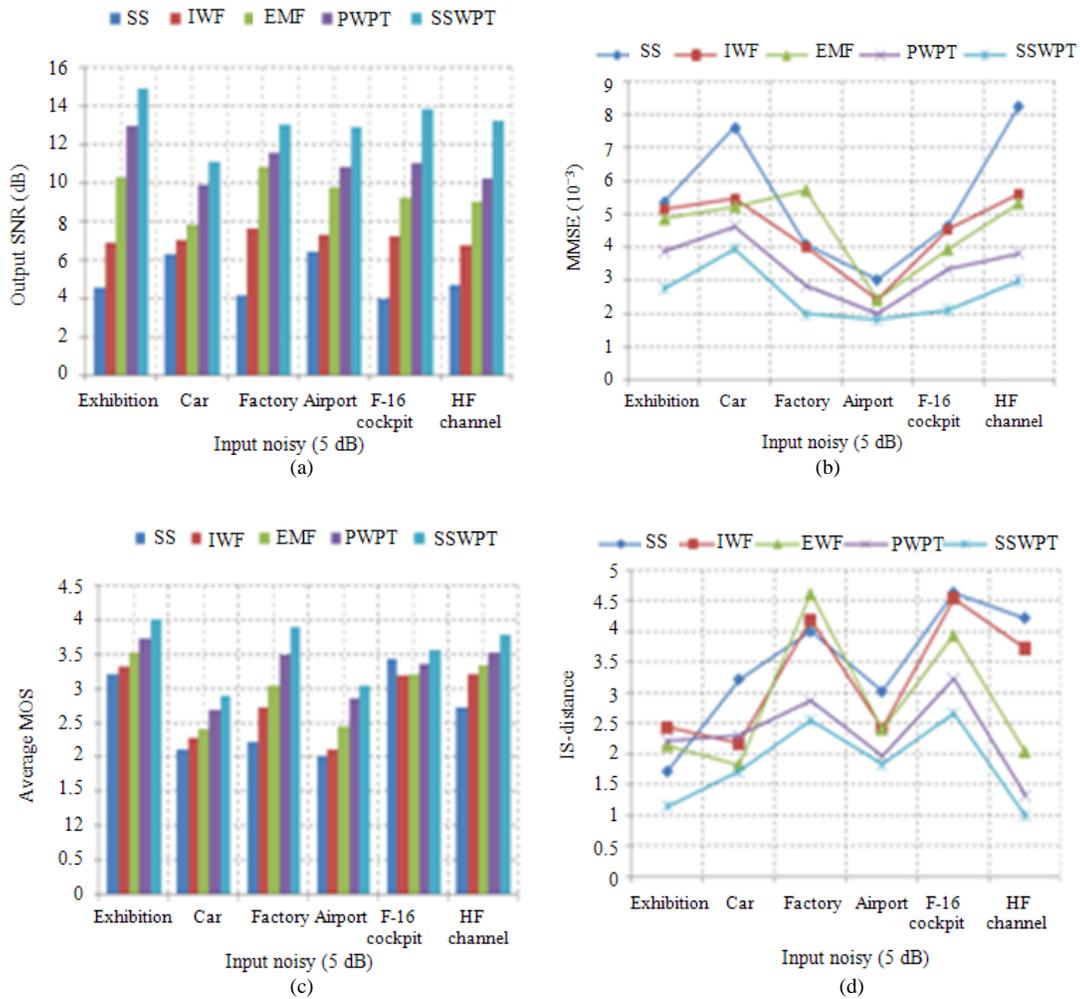


Fig. 10: Comparison of average performance metrics for various noises for different methods at 5dB SNR levels (a) Output SNR (b) average MOS (c) MMSE (d) IS-Distance

EMF and IWF is competitive to each other. Figure 10(c) indicates the MMSE performance for the noises as discussed in Fig. 10(a). It is seen that the enhancement of the proposed method is ahead of existing. The act of IWF and SS is competitive in airport noise. Figure 10(d) shows the spectral distance measure for the same six different types of noises as considered above. The obtained results indicate that SSWPT shows less spectral distance between enhanced and clean speech when compared to existing.

CONCLUSION

In this study the significance of combined spectral subtraction with wavelet packet based thresholding for enhancing the speech from the background noise is

demonstrated. Spectral subtraction uses dual band approach with adaptive noise estimator, that is in each band subtraction of the noise spectrum estimate is made to reduce the noise initially. To mask the effect of musical noise a scaled version of noisy spectrum is added and further it is processed in wave packet domain to get better speech quality. The performance of SSWPT is analyzed for different thresholding. Improved speech quality is obtained by the proposed SSWPT with modified thresholding.

The performance of the proposed SSWPT method can be increased irrespective of type of degradation by adaptively updating the thresholds or by combining with existing voice activity detection features. In this study no explicit study is made for processing unvoiced

regions of degraded speech. Therefore a rigorous analysis of unvoiced regions can be done in this method. In particular, the method has a need to develop for (i) identification of unvoiced sounds (ii) identification of speech specific spectral features of unvoiced sounds. The proposed wavelet based method need to be developed when the speech contains all three types of degradations (reverberation, additive noise and multi-speaker speech). In practical conditions, methods can be developed to identify the type of degradation and also level of degradation.

REFERANCES

- Ali, A.D., P.D. Swami and J. Singhai, 2010. Modified curvelet thresholding algorithm for image denoising. *J. Comput. Sci.*, 6: 18-23. DOI: 10.3844/jcssp.2010.18.23
- Ayache, M., M. Khalil and F. Tranquart, 2010. DWT to classify automatically the placental tissues development: Neural network approach. *J. Comput. Sci.*, 6: 634-640. DOI: 10.3844/jcssp.2010.634.640
- Boubakir, C. and D. Berkani, 2010. Speech enhancement using minimum mean-square error amplitude estimators under normal and generalized gamma distribution. *J. Comput. Sci.*, 6: 700-705. DOI: 10.3844/jcssp.2010.700.705
- Chomphan, S., 2010. Multi-Pulse based code excited linear predictive speech coder with fine granularity scalability for tonal language. *J. Comput. Sci.*, 6: 1288-1292. DOI: 10.3844/jcssp.2010.1288.1292
- Chavan, D.M.S., M.M.N. Chavan and D.M.S. Gaikwad, 2010. Studies on implementation of wavelet for denoising speech signal. *Int. J. Comput. Appl.*, 3: 1-7. DOI: 10.5120/711-1001
- Chen, S.H. and J.F. Wang, 2004. Speech enhancement using perceptual wavelet packet decomposition and teager energy operator. *J. VLSI Sign. Process.*, 36: 125-139. DOI: 10.1023/B:VLSI.0000015092.19005.62
- El-Leithy, S.T. and W.M. Sheta, 2009. Wavelet-based geometry coding for three dimensional mesh using space frequency quantization. *J. Comput. Sci.*, 5: 564-572.
- Ghanbari, Y. and M.R. Karami-Mollaei, 2006. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Commun.*, 48: 927-940. DOI: 10.1016/j.specom.2005.12.002
- Helmy, A.K. and G.S. El-Taweel, 2010. Neural network change detection model for satellite images using textural and spectral characteristics. *Am. J. Eng. Applied Sci.*, 3: 604-610. DOI: 10.3844/ajeassp.2010.604.610
- Hu, Y. and P.C. Loizou, 2004. Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE Trans. Speech Audio Process.*, 12: 59-67. DOI: 10.1109/TSA.2003.819949
- Hu, Y. and P.C. Loizou 2007. A comparative intelligibility study of single-microphone noise reduction algorithms. *J. Acoust. Soc. Am.*, 122: 1777-1786.
- Jiang, P., Q. Huang, Y. Kong and L. Chai, 2006. Research on a Denoising Method Based on Wavelet Packet Shrinkage for Pulp. *IEEE Proceedings of the 1st International Multi-Symposium on Computer and Computational Sciences*, June 20-24, IEEE Xplore, Hanzhou, Zhejiang, pp: 220-225. DOI: 10.1109/IMSCCS.2006.115
- Johnson, M.T., X. Yuan and Y. Ren, 2007. Speech signal enhancement through adaptive wavelet thresholding. *Speech Commun.*, 49: 123-133. DOI: 10.1016/j.specom.2006.12.002
- O'Shaughnessy, D., 2005. *Speech Communications Human and Machine*. 2nd Edn., Universities Press, India, ISBN-10: 817371374X, pp: 576.
- Rangachari, S. and P. Loizou, 2006. A noise-estimation algorithm for highly non-stationary environments. *Speech Commun.*, 48: 220-231. DOI: 10.1016/j.specom.2005.08.005
- Ramadan, Z.M., 2008. A three-microphone adaptive noise canceller for minimizing reverberation and signal distortion. *Am. J. Applied Sci.*, 5: 320-327. DOI: 10.3844/ajassp.2008.320.327
- Sumithra, M.G.A., K. Thanuskodi and M.R.C. Anitha, 2009. Modified time adaptive wavelet based approach for enhancing speech from adverse noisy environments. *ICGST Int. J. Digital Signal Proc.*, 9: 33-40.
- Stark, B. and B.D. Barkana, 2010. Acoustic echo cancellation: dual architecture implementation. *J. Comput. Sci.*, 6: 101-106. DOI: 10.3844/jcssp.2010.101.106
- Wang, K.C., 2010. Wavelet-based speech enhancement using time-frequency adaptation. *EURASIP J. Adv. Signal Process.*, 2009: 1-8. DOI: 10.1155/2009/924135