ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

Speech Enhancement in terms of Objective Quality Measures Based on Wavelet Hybrid Thresholding the Multitaper Spectrum

S.China Venkateswarlu[1], A.SubbaRami Reddy[2], K.Satya Prasad[3]

Associate Professor, Dept. of Elecrical & Computing, ASTU- Research Scholar, JNTUK-Kakinada, A.P, India[1]

Principal, SKIT, Srikalahasti, Chittoor Dt, A.P, India[2]

Ex-Rector & Professor, Dept. Of ECE JNTUK, Kakinada, A.P, India[3]

ABSTRACT: In this paper, Modified Improved Arctan Thresholding and two Hybrid Thresholding schemes (TH-I and TH-II) have been proposed and successfully used in Wavelet Thresholding the Multitaper Spectrum for Speech Enhancement. The two Hybrid Thresholding schemes are formulated by incorporating the modified arctan function and exponential function suitably in Hard and Soft thresholding methods. To address the “musical noise” encountered in most of the frequency domain speech enhancement algorithms, Yi. Hu and P.C. Loizou have proposed the Wavelet Thresholding the Multitaper Spectrum Speech Enhancement with Hard and Soft Thresholding schemes. This paper investigates further improvement of Speech quality in terms of six objective quality measures using Discrete Wavelet Transform Thresholding the Multitaper Power Spectrum with the above proposed Thresholding methods. The performance of the new Thresholding methods is compared with the other thresholding methods. It is observed that, further improvement is observed with the proposed schemes when applied to noisy speech signals of low SNR (0dB) conditions. The results showed that, the use of Multitaper Spectrum estimation combined with the proposed Thresholding procedures yields better quality in terms of six objective quality measures. This Algorithm incorporates the noisy speech signal divided in to overlapping frames and each frame is subject to the Multitaper Power Spectrum estimation using either SLEPIAN(Discrete Prolate Spheroidal Sequences) or SINE Tapers and applied to the Wavelet Thresholding the Multitaper Spectrum based Speech Enhancement Algorithm and the enhanced speech is reconstructed in its time domain. Analysis is done using Daubechies and Symlet wavelets with different real world noisy environments as well as white Gaussian Noise. Six Objective quality measures are considered in this study to test the performance of the algorithm for enhanced speech quality and compared. The proposed Thresholding methods perform better than hard, soft, Modified sigmoid and modified improved thresholding methods for Wavelet Thresholding the Multitaper Spectrum based Speech Enhancement.

KEYWORDS: Multitaper method, power spectrum estimation, wavelet hybrid thresholding, modified arctan function, speech enhancement.

I. INTRODUCTION

In Speech Processing, Speech enhancement is one of the most important fields and finds many applications such as mobile phones, teleconferencing systems, speech recognition and hearing aids. Speech signals from the uncontrolled environment may contain degradation components along with required speech components like background noise, speech from other speakers etc and decrease the intelligibility and the quality of the Speech. Reducing or suppressing such background noise and improving the perceptual quality and intelligibility of a speech without disturbing the speech signal quality is the main aim of speech enhancement to improve the quality and intelligibility of such degraded speech signal [1, 2]. The processed speech signals are supposed to be more comfort for listening and also should give better performance in tasks like automatic speech and speaker recognition [1]. During the last decades, several algorithms are proposed in the literature for speech enhancement such as spectral subtraction methods, MMSE methods, Weiner algorithm, Subspace methods, Hidden Markov Modelling, wavelet-based methods etc., have been proposed to solve this problem. Although the above speech enhancement algorithms improve speech quality, they suffer from an annoying artefact called “musical noise” [3-5]. Musical noise is caused by randomly spaced spectral peaks that come and go in each frame, and occur at random frequencies. The randomly spaced peaks are due to the inaccurate and large-variance estimates of the spectra of the noise and noisy signals, typically computed using periodogram type methods. Several methods have been proposed to reduce musical noise [6-9].

Yi. Hu and P.C. Loizou, in their paper [5], presented a different approach using the multitaper method proposed by Thomson [10] for power spectrum estimation. The multitaper method was shown in [10] to have good bias and variance properties. To further refine the spectral estimate, further they carried on wavelet thresholding the log multitaper spectra. Unlike others who wavelet denoised the time-domain signal (e.g., [11]), they wavelet denoise the speech spectrum. It should be pointed out that they have not used the wavelet denoising techniques to remove the noise, but rather to get better (lower variance) spectral estimates using the Hard and Soft thresholding techniques. In this paper an attempt has been made to propose new Thresholding schemes which are formulated by incorporating the Modified Arctan function in Hard and Soft, thresholding methods. The approach is similar to the two Hybrid Thresholding schemes reported by the authors in the recent paper [12] except the use of Wavelet Thresholding the Multitaper Spectrum for Speech Enhancement. The two Hybrid Thresholding schemes are formulated by incorporating the modified arctan function union with exponential function [13] and sigmoid function [14] union with the above modified arctan function in place of Hard and Soft thresholding methods [15-16]. The proposed new schemes are Modified Improved Arctan function Thresholding (MIAT), Modified Arctan function Thresholding (MAT) and two Hybrid Thresholding schemes (TH-I and TH-II) are used in Wavelet Thresholding the Multitaper Spectrum for Speech Enhancement. The details are presented in Section-III. This paper investigates further improvement of Speech quality in terms of six objective quality measures [17-19] using Discrete Wavelet Transform Thresholding the Multitaper Power Spectrum with the above proposed two Hybrid Thresholding methods based on arctan function. The performance of the new Hybrid methods is compared with the other thresholding methods. It is observed that the new proposed scheme yields further improvement when applied to noisy speech signals with low SNR (0dB) conditions. The results showed that, the use of Multitaper Spectrum estimation combined with wavelet Hybrid Thresholding suppressed the musical noise and yielded better quality than the existing hard, soft, improved and modified improved Thresholding methods. This Algorithm incorporates the noisy speech signal divided in to overlapping frames and each frame is subject to the Multitaper Power Spectrum estimation using either SLEPIAN (Discrete Prolate Spheroidal Sequences) or SINE Tapers and applied to the Wavelet Thresholding the Multitaper Spectrum based Speech Enhancement Algorithm suggested [5] and the enhanced speech is reconstructed in its time domain.
Analysis is done using Daubechies and Symlet wavelets with different real world noisy environments using Noizeus database [20] as well as white Gaussian Noise. Six Objective quality measures [17-19] are considered in this study to test the performance of the algorithm for enhanced speech quality and compared. Further improvement is reported with the proposed Hybrid Thresholding methods for Wavelet Thresholding the Multitaper Spectrum based Speech Enhancement than the use of hard, soft, Improved and modified improved thresholding methods.
This paper is organized as follows. Section II provides background information on Multitaper Spectrum Estimation, and Section III presents the proposed Hybrid Thresholding approach using the arctan function. The Scheme implementation details are presented in Section IV, the results are presented in Section V, and the conclusions are given in Section VI.

II. MULTITAPER SPECTRUM ESTIMATION

Direct spectrum estimation based on the use of Hamming windowing is the most commonly used power spectrum estimator for speech enhancement. Although windowing reduces the bias, it does not reduce the variance of the spectral estimate [21]. The concept behind the multitaper spectrum estimator [5, 10, 22] is to reduce this variance by computing a small number (L) of direct spectrum estimators each with a different taper (window), and then average the L spectral estimates. The idea is similar to the Welch's method of modified periodogram [21]. If the set of L tapers are chosen to be pair wise orthogonal and properly designed to prevent leakage then the resulting multitaper spectral estimator will be superior to the periodogram in terms of reduced bias and variance. At best, the variance of the multitaper estimate will be smaller than the variance of each spectral estimate by a factor of 1/L.

The multitaper spectrum estimator is given by

$$\sum_m a_k(m) a_j(m) = 0 \tag{1}$$

With

$$\hat{S}_k^{mt}(w) = \left| \sum_{m=0}^{N-1} a_k(m) x(m) e^{-jwm} \right|^2 \qquad (2)$$

Where N is the data length and a_k is the k[th] data taper used for the spectral estimate $\hat{S}_k^{mt}(.)$ which is also called the kth eigen spectrum. These tapers are chosen to be orthonormal, i.e., $\sum_m a_k(m) a_j(m) = 0$ for k$\neq$j and equal to 1 for k=j. A good set of L orthogonal data tapers with good leakage properties are given by the Slepian or discrete prolate spheroidal sequences (dpss) which are a function of a prescribed mainlobe width W. The number of tapers L is chosen to be less than 2NW, where W is expressed in units of normalized frequency, i.e.,0<W<1/2 [10]. The Slepian sequences are the unique orthogonal sequences which maximize the spectral concentration of the window mainlobe within [-W,W]. Other taper sequences were also proposed that minimize instead the local bias of the spectral window. In particular, Riedel and Sidorenko [23] proposed the sine tapers given by

$$a_k(m) = \sqrt{\frac{1}{N+1}} \sin\left(\frac{\pi(k+1)m}{N+1}\right), \qquad m = 0,1,2,.......N\text{-}1 \qquad (3)$$

The sine tapers were shown in [23] to produce smaller local bias than the Slepian tapers, with roughly the same spectral concentration. In this paper, we adopted both the tapers and compared.

III. THE WAVELET THRESHOLDING THE MULTITAPER SPECTRUM

It was already reported in the literature that wavelet thresholding techniques can be used to refine the spectral estimate and produce a smooth estimate of the logarithm of the spectrum [24-26]. The basic idea behind these techniques is to represent the log periodogram as "signal" plus the "noise," where the signal is the true spectrum and "noise" is the estimation error [27]. If the "noise" is Gaussian, then standard wavelet shrinkage techniques can be used to eliminate the "noise", by employing level-independent "universal" thresholds [28] for better spectral estimates. It was shown in [22] that if the eigen spectra defined in (2) are assumed to be uncorrelated, the ratio of the estimated multitaper spectrum $\hat{S}_k^{mt}(.)$ and the true power spectrum S(w) conforms to a chi-square distribution with 2L degrees of freedom, i.e.,

$$v(w) = \frac{\hat{S}_k^{mt}(.)}{S(w)} \sim \frac{X_{2L}^2}{2L}, \qquad 0 < w < \pi \qquad (4)$$

Taking the log of both sides, we get

$$\log \hat{S}_k^{mt}(.) = \log S(w) + \log v(w) \qquad (5)$$

Hence, the log of the multitaper spectrum can be represented as the sum of the true log spectrum plus a noise term, which is log X^2 distributed. If L is at least 5, the distribution of v(w) will be very close to a normal distribution with mean ϕ(L) - log L and variance ϕ'(L), where ϕ(L) and ϕ'(L) denote the digamma and trigamma functions respectively [5]. This means that for all ω (except near ω=0 and π) the random variable η(ω)

$$\eta(w) = \log v(w) - \phi(L) + \log L \qquad (6)$$

will be approximately Gaussian with zero mean and known variance $\sigma_\eta^2 = \phi'(L)$. If $Z(w)$ is defined as

$$Z(w) = \log \hat{S}_k^{mt}(w) - \phi(L) + \log L \qquad (7)$$

then

$$Z(w) = \log S(w) + \eta(w) \qquad (8)$$

So, the log multitaper power spectrum plus a constant ($\log L - \phi(L)$) can be written as the true log power spectrum plus a nearly Gaussian noise η(w) with zero mean and known variance σ_η^2 [25]. The model in (8) is well suited for

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

wavelet denoising techniques [28-31] for eliminating the noise and obtaining a better estimate of the log spectrum. The idea behind refining the multitaper spectrum by wavelet thresholding can be summarized in four steps [5]. 1) **O**btain the multitaper spectrum using (1) and (3), and calculate $Z(w)$ using (7). 2) Apply a standard, periodic Discrete Wavelet Transform (DWT) out to level q_0 to $Z(w)$ in order to get the empirical DWT coefficients $z_{j,k}$ at each level j, where q_0 is specified in advance [32]. 3) Apply a thresholding procedure to $z_{j,k}$ (the scaling coefficients are kept intact). 4) Apply the inverse DWT to the thresholded wavelet coefficients and obtain the refined log spectrum.
One of the key steps in the above process is the thresholding procedure, which is critical to the performance of the algorithm [5]. The details about the thresholding techniques will be given in Section III-B.

IV. SPEECH ENHANCEMENT BY WAVELET THRESHOLDING THE MULTITAPER SPECTRUM

In this section, the STSA estimator, derived by Yi. Hu and P.C. Loizou [5] is presented for clarity.

A. Proposed STSA Estimator

Here it is assumed that the noise signal is additive and uncorrelated with the speech signal, and is given by y=x+n
y, x and n are N-dimensional noisy Speech, clean Speech and noise vectors respectively. If F is the N-point discrete Fourier Transform matrix then the Fourier transform of the noisy speech vector y can be written as

$$Y(w) = F^H.y = F^H.x + F^H.n = X(w) + N(w)$$

where $X(w)$ and $N(w)$ are the N×1 vectors containing the spectral components of the clean speech vector x and the noise vector n, respectively. The linear estimator of the clean speech Spectrum is given by

$$\hat{X}(w) = G.Y(w)$$

Where G is a N×N matrix. The error signal of this estimation is expressed by

$$\varepsilon(w) = \hat{X}(w) - X(w) = \varepsilon_x(w) + \varepsilon_n(w)$$

Where $\varepsilon_x(w)$ represents the speech distortion in the frequency domain and $\varepsilon_n(w)$ represents the residual noise in the frequency domain and can be expressed as

$$\varepsilon_x(w) = (G - I).X(w)$$

$$\varepsilon_n(w) = G.N(w)$$

In frequency domain, the energy of the speech distortion and the energy of the residual noise respectively can be given by

$$\varepsilon_x^2(w) = E\left(\varepsilon_x^H(w).\varepsilon_x(w)\right)$$

$$\varepsilon_n^2(w) = E\left(\varepsilon_n^H(w).\varepsilon_n(w)\right)$$

The optimal linear estimator can be obtained by solving the following constrained optimization problem:

$$\min_G \varepsilon_x^2(w), \qquad \text{Subject to: } \frac{1}{N}\varepsilon_n^2(w) \le c \qquad (9)$$

where c is a positive number. It was established that [33] the optimal G satisfies the following equation:

$$G\left(F^H.R_x.F + \mu F^H.R_n.F\right) = F^H.R_x.F \qquad (10)$$

Where μ is the Lagrange multiplier. The above equation can be simplified by assuming that each frequency component is modified individually by a gain, that is, if it is assumed that G is a diagonal matrix. The matrices $F^H.R_x.F$ and $F^H.R_n.F$ are asymptotically diagonal [34] (assuming that R_x and R_n are Toeplitz) and the diagonal elements of $F^H.R_x.F$ and $F^H.R_n.F$ are the power spectrum components $S_x(w)$ and $S_n(w)$ of the clean speech

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

vector x and the noise vector n, respectively. Denoting the k th diagonal element of G by $g(k)$, then for large N, (10) can be rewritten as

$$g(k)=\frac{S_x(k)}{S_x(k)+\mu S_n(k)}=\frac{\gamma_{prio}(k)}{\gamma_{prio}(k)+\mu} \qquad (11)$$

Here $\gamma_{prio}(k)$ is the *a priori SNR* at frequency w_k and is defined as

$$\gamma_{prio}(k)=S_x(w)/S_n(w)$$

The Lagrange multiplier $\mu(\mu \geq 0)$ controls the trade-off between speech distortion and residual noise [34]. This can be illustrated by noting that μ is related to $\varepsilon_x^2(k)$ and $\varepsilon_n^2(k)$ by

$$\varepsilon_x^2(k)=\sum_k (1-g(k))^2 S_x(k)=\sum_k \left(\frac{\mu}{\gamma_{prio}(k)+\mu}\right)^2 S_x(k)$$

$$\varepsilon_n^2(k)=\sum_k g^2(k) S_n(k)=\sum_k \left(\frac{\gamma_{prio}(k)}{\gamma_{prio}(k)+\mu}\right)^2 S_n(k)$$

A large μ, would produce more speech distortion with less residual noise. Similarly, a small μ would produce smaller amount of speech distortion containing more residual noise. Ideally, ideally one has to minimize the speech distortion in speech-dominated frames since the speech signal will mask the noise in those frames. Similarly, one would like to reduce the residual noise in noise-dominated frames. To accomplish that, one can make the value of dependent on the estimated *a priori SNR*. Hence the following equation has been chosen [5] for estimating μ:

$$\mu=\mu_0-\frac{SNR_{dB}}{s} \qquad (12)$$

Where μ_0 and s are constants chosen experimentally, and SNR_{dB} is given by

$$SNR_{dB}=10\log_{10} SNR$$

The power spectrum $S_x(w)$ in (11) of the clean speech signal is not available, but in practice it can be estimated as:

$$\hat{S}_x(w)=S_y(w)-\hat{S}_n(w) \qquad (13)$$

Where $\hat{S}_n(w)$ denotes the estimate of the noise spectrum obtained during speech-absent frames. As the estimate of the $\gamma_{prio}(k)$ is crucial for eliminating musical noise, it is considered two different methods for obtaining a good estimate of $\gamma_{prio}(k)$.

In the first method, the ratio of the multitaper spectra $\hat{S}_x^{mt}(w)/\hat{S}_n^{mt}(w)$, and wavelet threshold the log of the ratio of the two spectra to get an estimate of $\gamma_{prio}(k)$. It can be proved [5] that the log *a priori SNR* estimate, based on multitaper spectra is denoted as $\gamma_{prio}^{mt}(k)$, can be modelled as the true log *a priori SNR* plus a Gaussian distributed noise $\xi(k)$, i.e.,

$$\gamma_{prio}^{mt}(k)=\log\gamma_{prio}(k)+\xi(k) \qquad (14)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

Where $\xi(k)$ is approximately Gaussian distributed with zero mean and known variance $2\sigma_\eta^2$. Because of the nature of $\xi(k)$, wavelet denoising techniques can be used to eliminate $\xi(k)$.

The second method is based on the assumption that a good estimate of the *a priori SNR*, can be obtained using a good low variance spectral estimate of $\hat{S}_x(w)$ and $\hat{S}_n(w)$.It considered first obtaining the multitaper spectral estimates of $S_y(w)$ and $\hat{S}_n(w)$ and then wavelet thresholding the log of those estimates individually to obtain $\hat{S}_x(w)$. The refined spectrum of $\hat{S}_n(w)$ and the wavelet thresholded estimate of $\hat{S}_n(w)$ are then used to obtain a better estimate of the *a priori SNR.* The above Wiener-type STSA estimator given in (11) is not new, although it was derived differently. What is new, however, is the finding that the log *a priori SNR*, if estimated using multitaper spectra, can be modelled as the true log *a priori SNR* plus a Gaussian distributed noise.

B. Wavelet Thresholding Techniques

The purpose of the thresholding procedure is to eliminate or suppress small value wavelet coefficients which mainly represent the noise content therefore the choice of threshold levels and thresholding techniques are critical to the performance of wavelet denoising techniques. Several methods have been proposed in the literature for thresholding the wavelet coefficients [28-31]. The thresholding techniques examined in this paper are described below. Let $\{z_{j,k}\}$ be the wavelet coefficients of $Z(w)$ in (8), let $\{s_{j,k}\}$ be the wavelet coefficients of $\log S(w)$ and let $n_{j,k}$ be the wavelet coefficients of $\eta(w)$. Then by the linearity of the discrete wavelet transform, (8) is transformed to

$$z_{j,k} = s_{j,k} + n_{j,k} \qquad (15)$$

Where the subscript j indicates the j th scale, and the subscript k indicates the k th wavelet coefficient. Since the noise $\eta(w)$ is nearly Gaussian, the standard thresholding functions used in the wavelet based enhancement systems are hard and soft thresholding functions [29] which we review before introducing a new thresholding function that offers improved performance for speech signal. Hard thresholding sets to zero any element whose absolute value is lower than the threshold. In Soft thresholding, the elements whose absolute values are lower than the threshold are first set to zero and then shrink the nonzero coefficients towards zero.

In these techniques, T is the threshold value and δ is the thresholding function.

i. The Hard Thresholding (HT):

The hard thresholding function is defined for threshold T as:

$$\delta_H(z,T) = \begin{cases} z, & \text{if} \quad |z| \ge T \\ 0, & \text{otherwise} \end{cases} \qquad (16)$$

ii. The Soft Thresholding (ST):

The soft thresholding function is defined as:

$$\delta_S(z,T) = \begin{cases} z-T, & \text{if} \quad z \ge T \\ 0, & \text{if} \quad |z| < T \\ z+T, & \text{if} \quad z \le -T \end{cases} \qquad (17)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

iii. The Modified Improved Thresholding (MIT):

The Modified Improved thresholding [13] is proposed by Asser Ghanbari and Mohammad Reza Karami and can be used like a hard thresholding function for the wavelet coefficients absolute value greater than threshold value and is like an exponential function for the wavelet coefficients absolute value less than threshold value and is defined [13] as:

$$\delta_{MI}(z,T)=\begin{cases} z=z & \text{if} \quad |z|\geq \mathrm{T} \\ sign(z)\dfrac{T\{\exp(\gamma_1\frac{|z|}{T})-1\}}{\{\exp(\gamma_1)-1\}} & \text{if} \quad |z|<\mathrm{T} \end{cases} \tag{18}$$

Here, the factor γ_1 is important and for this work γ_1=3 is used in order to have better performance [13].

iv. The Modified Sigmoid function Thresholding (MST):

Modified Sigmoid function thresholding proposed by Ting-Hua.Yi et al. [14] attempts to address the deficiency of hard and soft thresholding denoising methods and is given as:

$$\delta_{MS}(z,T)=\begin{cases} (z-T)-T\left[\dfrac{2}{1+e^{\beta\frac{(z-T)}{T}}}-1\right] & \text{if} \quad z\geq T \\ 0 & \text{if} \quad |z|<T \\ (z+T)-T\left[\dfrac{2}{1+e^{\beta\frac{(z+T)}{T}}}-1\right] & \text{if} \quad z\leq -T \end{cases} \qquad \beta\in R^{+} \tag{19}$$

In this function, the factor β is important and for this work β =4.5 is used in order to have better performance [14].

v. ARCTAN FUNCTION BASED THRESHOLDING TECHNIQUE

An important part of a wavelet-based speech enhancement system is the thresholding function. Ideal thresholding functions retain or shrink only wavelet coefficients exceeding a threshold value T. However, comparing both hard- and soft-shrinking schemes, it can be seen that the hard thresholding exhibits some discontinuities at ±T and may be unstable or more sensitive to small changes in the data. On the other hand, in soft thresholding the wavelet coefficients are reduced by a quantity equal to the threshold value which will induce the deviation when the filtered signal is reconstructed by the inverse WT. To overcome this shortcoming, an alternative procedure based on the sigmoid function is proposed [14]. This procedure provides a compromise between the advantages and drawbacks of hard- and soft-thresholding.

In this proposed work the sigmoid function in [14] is replaced with modified arctan function (20) defined as:

$$z=\frac{2T}{\pi}Tan^{-1}(\beta x), \qquad \beta\in R^{+} \tag{20}$$

The arctan function is real-valued and differentiable. The degree of approximation of the arctan function to the Signum function can be adjusted by regulating the value of β. Here the modification to the arctan function is introduced in the magnitude and argument terms by 2T/π and β *x* respectively. The modified arctan function described in (20) produces an output z over the range extended from −T to +T. Here the factor 2/π is introduced in the magnitude for normalization

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

of the arctan function to T value. Figure 2 shows the typical plot of the arctan function with β equal to 0.05, 0.1, 0.5, 1, 2, 5, 10 and 100.

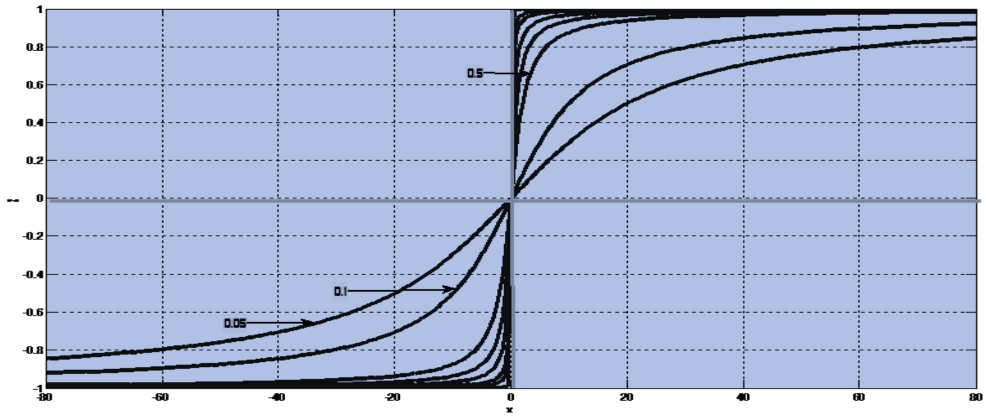


Figure 2: Plot of the modified arctan function.

a) The Modified Improved Arctan function Thresholding (MIAT):

The modified arctan function can be used as thresholding scheme in a similar fashion as proposed in [13] by replacing the exponential function with the modified arctan function defined in (20) and can be used like a hard thresholding function for the wavelet coefficients absolute value greater than threshold value and for the coefficients with absolute value less than the threshold value were mapped on an modified arctan function instead of setting to zero and is defined as:

$$\delta_{MIA}(z,T)=\begin{cases} z=z & \text{if} \quad |z|\geq \mathrm{T} \\ \dfrac{2T}{\pi}Tan^{-1}(\beta z), & \text{if} \quad |z|<\mathrm{T}, \quad \text{and} \quad \beta\in R^{+} \end{cases} \tag{21}$$

The input-output characteristics of Modified Improved Arctan function thresholding is shown in the figure.3 for β =0.8.

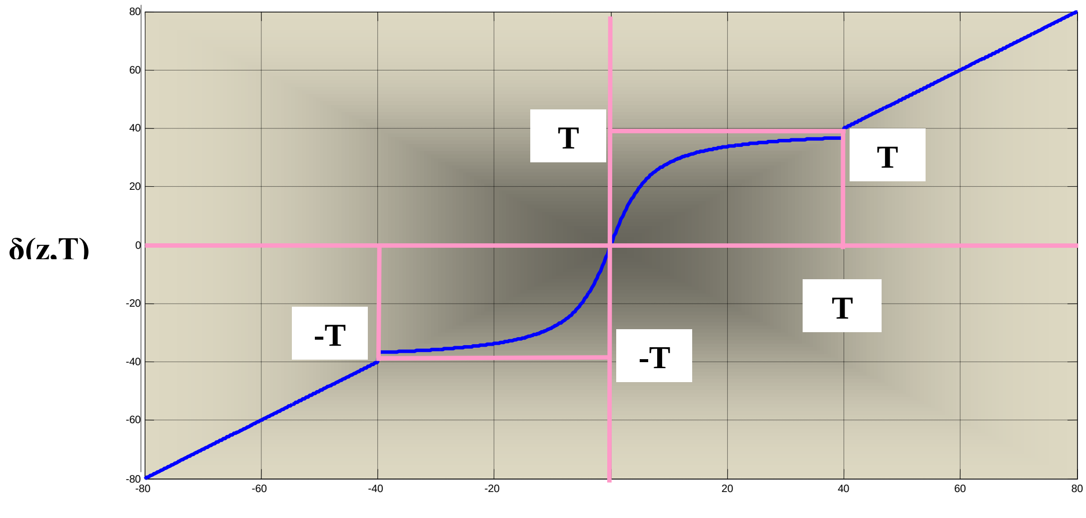


Figure 3: The input-output characteristics of thresholding function.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

In the above, the factor β plays an important role and to achieve better results, in this work we selected β =1.2 based on the overall performance achieved through the six objective measures. Phonetically balanced clean speech signals have been taken from a speech corpus called “NOIZEUS” [20] and White Gaussian noise is added to these clean speech signals at 0dB, 5dB, 10dB, and 15dB SNR levels to generate Noise corrupt speech signals. Wavelet Thresholding Multitaper speech enhancement algorithm which makes use the Arctan function with different β values ranging from 0 to 1.8 in steps of 0.2 is applied to the noise corrupt speech signals. For each value of β the performance of the enhanced signal is analyzed by using six speech quality objective measures LLR, WSS, CEPSTRAL, SEG-SNR, FWSEG-SNR, and SNR [17-18, 35-37] for speech enhancement and are shown in the fig.4. All the measures are computed by segmenting the sentences using 32-ms duration Hamming window with 75% overlap between adjacent frames. A tenth order LPC analysis was used in the computation of LPC- based objective measure LLR. The wavelet we used was “db8”, but for most of the wavelets, the results don’t have considerable changes. From the simulation results it is verified that, β is selected in between 0.8 and 1.2 to achieve optimum results.

b) The Modified Arctan function Thresholding (MAT):

In this work we proposed a similar type thresholding function presented in [14] based on modified arctan function. This new scheme is termed as Modified Arctan function thresholding scheme and can be defined as:

$$\delta_{MA}(z,T)=\begin{cases}(z-T)+\dfrac{2T}{\pi}Tan^{-1}\{\beta_2\dfrac{(z-T)}{T}\} & \text{if} \quad z\geq T\\ 0 & \text{if} \quad |z|<\text{T}\\ (z+T)+\dfrac{2T}{\pi}Tan^{-1}\{\beta_2\dfrac{(z+T)}{T}\} & \text{if} \quad z\leq -T\end{cases} \tag{22}$$

Figure.5 illustrates the comparison of soft, hard and modified arctan function based thresholding scheme (with T = 40 and β_2 equal to 0, 0.5, 1.0, 3.0, 5.0, 8.0, 15.0 and 30.0). From the figure.5, it is evident that the scheme sets the threshold function values in between the values obtained through hard and soft thresholding procedures and thus provides a compromise between the advantages and drawbacks of hard- and soft-thresholding schemes.

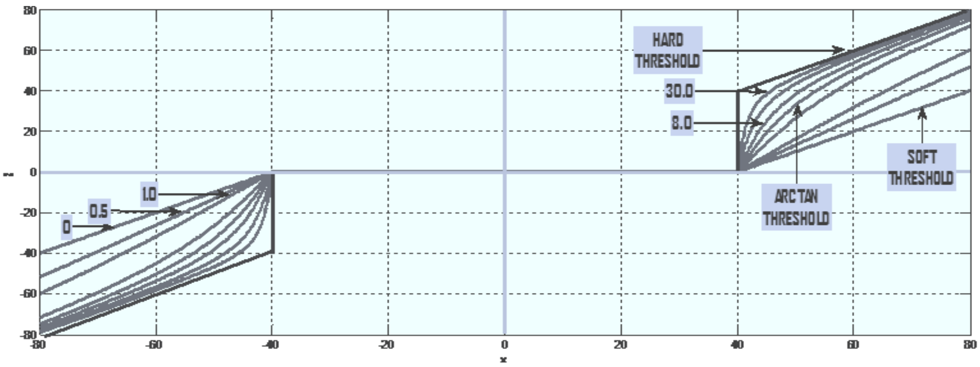


Figure 5: The comparison of soft, hard and arctan function based thresholding algorithm.

Based on the modified arctan function, this work also proposes two Hybrid Thresholding schemes TH-I and TH-II. The results shown that, the proposed Hybrid schemes overcome the drawbacks of both the Hard and Soft thresholding functions by providing a smooth transition to the wavelet coefficients with absolute value less than threshold value instead of setting the coefficients to zeros.

In the proposed Hybrid thresholding scheme, a modification is made to the hard and soft thresholding functions. In Hybrid Thresholding(TH-I), the wavelet coefficients with absolute value less than the threshold value were mapped on to an arctan function as described in (20) instead of setting to zero and the wavelet coefficients with absolute value

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

greater than the threshold value were mapped on to a modified arctan function as described in (22). Similarly in Hybrid Thresholding (TH-II), the wavelet coefficients with absolute value less than the threshold value were mapped on to an exponential function (18) instead of setting to zero and the wavelet coefficients with absolute value greater than the threshold value were mapped on to the same modified arctan function as described in (22). The Hybrid Thresholding schemes are defined as:

c). Hybrid Thresholding-I (TH-I) scheme is defined as:

$$\delta_{HTA}(z,T)=\begin{cases}(z-T)+\dfrac{2T}{\pi}Tan^{-1}\{\beta_2\dfrac{(z-T)}{T}\}z & \text{if} \quad z\geq T\\ \dfrac{2T}{\pi}Tan^{-1}(\beta_1 z) & \text{if} \quad |z|<\mathrm{T}\\ (z+T)+\dfrac{2T}{\pi}Tan^{-1}\{\beta_2\dfrac{(z+T)}{T}\} & \text{if} \quad z\leq -T\end{cases} \tag{23}$$

In the above, two important factors β_1 and β_2 which are real and positive constants and we determined the values for optimum results using Speech quality objective measures. For the wavelet coefficients with the absolute value greater than threshold value, it can be easily proved that the presented thresholding scheme is a kind of compromise between hard and soft thresholding, where the difference is due to the constants β_1 and β_2. If β_1=0 and β_2 = 0, soft thresholding can be considered, and if β_1=0 and $\beta_2 \to +\infty$, the equation corresponds to hard thresholding.

d). Hybrid Thresholding-II (TH-II) is defined as:

$$\delta_{HTE}(z,T)=\begin{cases}(z-T)+\dfrac{2T}{\pi}Tan^{-1}\{\gamma_2\dfrac{(z-T)}{T}\} & \text{if} \quad z\geq T\\ sign(z)\dfrac{T\{\exp(\gamma_1\frac{|z|}{T})-1\}}{\{\exp(\gamma_1)-1\}} & \text{if} \quad |z|<\mathrm{T}\\ (z+T)+\dfrac{2T}{\pi}Tan^{-1}\{\gamma_2\dfrac{(z+T)}{T}\} & \text{if} \quad z\leq -T\end{cases} \tag{24}$$

In this function, the factors γ_1 and γ_2 are real and positive constants and we determined the value for optimum results using Speech quality objective measures. In this scheme also it can be easily proved that the presented thresholding scheme is a kind of compromise between hard and soft thresholding for the wavelet coefficients with absolute value greater than the threshold value T and the constant γ_2 is selected properly.

The Speech Enhancement results in terms of speech objective quality measures obtained with the above the proposed Modified Improved Arctan Thresholding (21), Modified Arctan Thresholding (22), Hybrid Thresholding scheme-I (23) and Hybrid Thresholding scheme-II (24), are compared with the results obtained through Hard Thresholding (16), Soft Thresholding (17), Modified Improved Thresholding (18) and Sigmoid function Thresholding (19) schemes and are presented in the Table.1(a) and Table 1(b).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

C. Wavelet Threshold Methods

1) Universal Threshold Method:

The main issue in wavelet denoising is determining an appropriate threshold level T . For i.i.d. zero-mean, normally distributed noise with variance σ , Donoho and Johnstone [29] proposed the so-called universal threshold given by $T=\sigma\sqrt{2\log N}$. It is simple since it does not depend on the input data, but on the noise variance, and works well for uncorrelated noise. If the noise is stationary and colored, the variance of the noise wavelet coefficients will be different for each scale in the wavelet decomposition [31]. Consequently, scale-dependent thresholding was proposed to account for the different variances of the noise wavelet coefficients in each scale. Therefore the level-dependent variances of the noise wavelet coefficients were estimated according to [25].

$$\mathrm{var}(n_{j,k})=\sigma_j^2\equiv\frac{1}{N}\sum_{k=0}^{N-1}S(k)\left|H_j(k)\right|^2 \qquad (25)$$

Where $H_j(k)$ is the frequency response of the length N periodized wavelet filter of level j, and $S(k)$ is the Fourier transform of the autocorrelation function $r_{\eta\eta}$ of the noise $\eta(w)$ which is approximated by [25]

$$r_{\eta\eta}(i)=\begin{cases}\sigma_\eta^2\left(1-\frac{|i|}{L+1}\right), if\ |i|\le L+1\\ 0, otherwise\end{cases} \qquad (26)$$

Where L is the number of tapers, and σ_η^2 is the variance of $\eta(w)$.

In scale-dependent universal thresholding, the threshold T at each scale j is selected as $T=\sigma_j\sqrt{2\log N}$. The wavelet coefficients $z_{j,k}$ can be thresholded at each level j using either $\delta_H(.,.)$ or $\delta_S(.,.)$ etc.

$$\hat{z}_{j,k}=\begin{cases}\delta(z_{j,k},T), & \text{if} \quad 1\le j\le q_0\\ z_{j,k}, & \quad j>q_0\end{cases} \qquad (27)$$

where q_0 is, some specified coarse resolution level.

2) SURE Method: The universal thresholding method tends to use a high threshold level, and in many cases it oversmooths the noisy signal [30]. In this work also the authors observed that the universal thresholding method exhibits very inferior performance when applied to evaluate the speech quality objective measures. In [30], Donoho and Johnstone showed that the Stein's unbiased risk estimator (SURE) could be used as the unbiased estimate of the MSE for the soft-thresholding scheme. Johnstone and Silverman [31] later generalized this idea to the case of colored noise, and showed that the SURE method can also be used in the presence of correlated noise and the present work is confined only SURE threshold method.

The Stein's unbiased estimate of the risk for a specific threshold T and input signal $\mathbf{x}=\{x_i\}_{i=1}^N$ using the soft thresholding function is given by [31]

$$\hat{U}(x,T)=\hat{\sigma}^2N+\sum_{i=1}^{N}\left\{\min(x_i^2,T^2)-2\hat{\sigma}^2I(|x_i|\le T)\right\} \qquad (28)$$

Where $\hat{\sigma}^2$ is the variance of the noise and I is the indicator function ($I(.)=1$ if $|x_i|\le T$ and $I(.)=0$ if $|x_i|>T$). The SURE threshold is obtained as [30], [31]:

$$T=\arg\min_{0\le T\le\hat{\sigma}\sqrt{2\log N}}\hat{U}(x,T) \qquad (29)$$

For level-dependent thresholding, the noise variance $\hat{\sigma}_j^2$ for level j can be obtained using the median absolute deviation (*MAD*) from zero estimates [31]:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

$$\hat{\sigma}_j = \frac{MAD(z_{j,k})}{0.6745} \tag{30}$$

where 0.6745 is a normalization factor. The operator MAD picks out the median of the absolute values of all the wavelet coefficients $z_{j,k}$ at resolution level *j*. As the universal threshold method gives high threshold value [5], in this study the SURE method is considered only.

V. IMPLEMENTATION DETAILS

For each speech frame, the Scheme can be implemented in four steps [5].

Step 1) Compute the multitaper power spectrum S_y^{mt} of the noisy speech ***y*** using (1), and estimate the multitaper power spectrum S_x^{mt} of the clean speech signal ***x*** by:

$$S_x^{mt} = S_y^{mt} - S_n^{mt}$$

where S_n^{mt} is the multitaper power spectrum of the noise. S_n^{mt} can be obtained using noise samples collected during speech absent frames and update the noise activity appropriately. Here *L* is set to 5. Any negative elements obtained are floored as follows:

$$S_x^{mt}(w) = \begin{cases} S_y^{mt}(w) - S_n^{mt}(w), & \text{if} \quad S_y^{mt}(w) > S_n^{mt}(w) \\ \beta S_n^{mt}(w) & \text{if} \quad S_y^{mt} \le S_n^{mt}(w) \end{cases} \tag{31}$$

Where β is the spectral floor set to β =0.002.

Step 2) Estimate the *a priori SNR* using one of the two methods described in Section III-A. In this work the second method is followed. Accordingly, first compute $Z(w)$ using the relation

$$Z(w) = \log S_m^{mt}(w) - \phi(L) + \log L \tag{32}$$

and then apply the Discrete Wavelet Transform (DWT) of $Z(w)$ out to level q_0 to obtain the empirical DWT coefficients $z_{j,k}$ for each level *j* [in this paper, q_0 was set to 5 as per [5]. Eighth-order Daubechie's wavelets with least asymmetry and highest number of vanishing moments, for a given support, were used. Threshold the wavelet coefficients $z_{j,k}$ with one of the Eight thresholding techniques viz. Hard using (16), Soft using (17), Modified Improved using (18), Modified Sigmoid function using (19), Modified Improved arctan function using (21), Modified Arctan using (22), Hybrid Thresholding with Arctan function (TH-I) using (23), and Hybrid Thresholding with Exponential function (TH-II) using (24) described in Section III-B, and apply the inverse DWT to the thresholded wavelet coefficients to obtain the refined log spectrum, $\log S_y^{wmt}(w)$, of the noisy signal. Repeat the above procedure to obtain the refined log spectrum, $\log S_n^{wmt}(w)$, of the noise signal. The estimated power spectrum $S_x^{wmt}(w)$ of the clean speech signal can be estimated using (13). The *a priori SNR* $\gamma_{prio}(k)$ estimate for frequency w_k is computed as:

$$\gamma_{prio}(k) = \frac{S_x^{wmt}(w_k)}{S_n^{wmt}(w_k)} \tag{33}$$

Step 3) Compute the μ value in (11) according to the segmental SNR:

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

$$\mu = \begin{cases} \mu_0 - \dfrac{(SNR_{dB})}{s}, & -5 < SNR_{dB} < 20 \\ 1, & SNR_{dB} \geq 20 \\ \mu_{\max}, & SNR_{dB} \leq -5 \end{cases} \tag{34}$$

Where $\mu_{\max}$ is the maximum allowable value of μ, which was set to 10, $\mu_0 = (1 + 4\mu_{\max})/5$, $s = 25/(\mu_{\max} - 1)$, $SNR_{dB} = 10\log_{10} SNR$ and the SNR is computed as:

$$SNR = \frac{\sum_{i=0}^{N-1} S_x^{wmt}(i)}{\sum_{i=0}^{N-1} S_n^{wmt}(i)} \tag{35}$$

Step 4) Estimate the gain function *g(k)* for frequency component w_k using (11). Compute the enhanced spectrum as:

$$\hat{X}(w_k) = g(k).\hat{Y}(w_k) \tag{36}$$

Apply the inverse FFT of to obtain the enhanced speech signal.
The above estimator was applied to 32-ms duration frames of the noisy signal with 50% overlap between frames. The enhanced speech signal was combined using the overlap and add method. The entire scheme is represented using flow diagram shown in results.

VI. EXPERIMENTAL RESULTS

To study the performance of any algorithm, combinations of subjective and objective measures have to be carried on. Currently, the accurate method for evaluating speech quality is through subjective listening tests. But it is costly and time consuming. Hence, six Objective measures are chosen to evaluate the performance of the proposed schemes of Thresholding the Multitaper Spectra Speech enhancement. Phonetically balanced clean speech signals and real world noise corrupted signals at different SNR levels have been taken from a speech corpus called “NOIZEUS” data base [20]. White noise was added to the same clean speech signals at different SNR levels to obtain white noise signals which are also used in this work. These noisy speech examples are at different input SNRs equal to 0dB, 5dB, 10dB and 15dB.

Each noisy sentence was enhanced by Eight Thresholding methods: Hard Thresholding using (16), Soft Thresholding using (17), Modified Improved Thresholding using (18), Modified Sigmoid function Thresholding using (19), Modified Improved arctan function Thresholding using (21), Modified Arctan Thresholding using (22), Hybrid Thresholding with Arctan function (TH-I) Thresholding using (23), and Hybrid Thresholding with Exponential function (TH-II) Thresholding using (24) described in Section III-B are used are used to wavelet-threshold the multitaper spectra with SURE thresholding discussed in Section III-A. All the measures are tested for both DPSS and Sinusoidal tapers and the results presented in the respective Table.1(a) and Table.1(i).

A. Objective Measure Evaluation

P.Loizou has presented a correlation analysis of Objective Quality measures for evaluating speech enhancement algorithms [17]. In this paper six measures namely SNR, Segmental SNR(Seg-SNR), Log Likelihood Ratio(LLR), Weighted spectral slope distance(WSS), Frequency weighted segmental SNR (fwseg-SNR) and Cepstral Distance (Cep)[17-18,35-37] are selected for performance evaluation test, considering the fact that Fwseg-SNR, LLR, Cep and WSS have high correlation with overall speech quality. The correlation coefficients for these measures with speech quality are 0.84, 0.85, 0.79 and 0.64 respectively [17]. These objective measures also have good correlation with subjective scores. Although the correlation coefficient of Seg-SNR is 0.36, it is chosen as a time domain measure where as the above measures are of frequency domain.

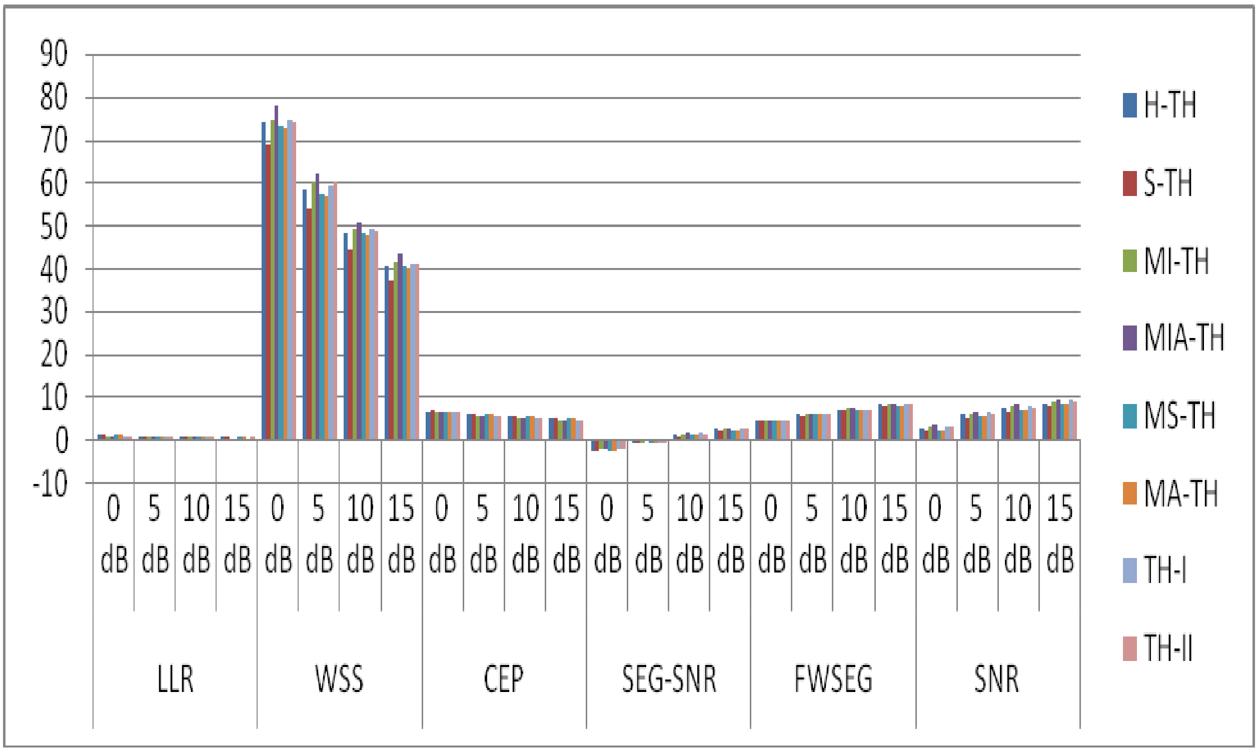
ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

The performance of the enhanced signal is analyzed by using six objective measures for speech enhancement. All the measures are computed by segmenting the sentences using 32-ms duration Hamming window with 75% overlap between adjacent frames. A tenth order LPC analysis was used in the computation of LPC- based objective measure LLR. Using both the DPSS and SINE tapers, the performance of the proposed Scheme is studied under three real world noise conditions namely “Airport noise” “Babble noise”, “Car noise” ,”Exhibition noise”, “Restaurant noise”, “Station noise”, “Street noise” and “Train noise” as well as “White Gaussian” noise at 0dB, 5dB, 10dB and 15dB input SNR values. The results are presented in below figure.



Figure(6): Graphical Comparative Performance of Thresholding Methods in terms of six Objective Quality Measures for Train Noise using DPSS Tapper..

In this work the Wavelet decomposition level q0 =5 is selected as there is no further advantage achieved for higher values of q0 [5] and the number of tapers L is selected in both DPSS and SINE tapers as L=5.

Best performance was obtained in all the six objective measures by the proposed Modified Improved arctan function Thresholding using (21) and Hybrid Thresholding with Arctan function (TH-I) using (23), based on multitaper spectra when compared to all the thresholding schemes presented in this work. Better performance is achieved with the Modified Arctan using (22) and Hybrid Thresholding with Exponential function (TH-II) using (24) when compared to Hard Thresholding using (16), Soft Thresholding using (17) when compared to Modified Sigmoid function using (19).

The Modified Improved Thresholding using (18) yields better results when compared to Hard Thresholding, Soft Thresholding and Modified Sigmoid function Thresholding.

The performance of the Modified Improved Thresholding is comparable to the performance of the proposed Hybrid Thresholding with Exponential function (TH-II) scheme. Another noteworthy observation is made in the present study when the SINE tapers are used and in the case of practical noises like Babble, Restaurant and Street, the performance of the Modified Improved arctan function Thresholding using (21) and Hybrid Thresholding with Arctan function (TH-I) using (23) are slightly inferior when compared with the Modified Improved Thresholding. Hence to achieve better quality improvement, selection of the Thresholding method and appropriate orthogonal tapers are important depending on the nature of the noise.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

Therefore this study demands for the design of new orthogonal tapers and appropriate Thresholding procedures to enhance fine details of the noisy speech signal. By Using both the DPSS and SINE tapers, the performance of the proposed Scheme is studied under three real world different noise conditions are observed. In this paper noise condition namely "Train noise" as well as "White Gaussian" noise at 0dB, 5dB, 10dB and 15dB input SNR values. The results are presented in Table.1 (a) and Table.1 (b).

Table.1(a): Comparative Performance of Thresholding Methods in terms of six Objective Quality Measures LLR, WSS, CEP, SEG-SNR, FWSEG and SNR Measures for real world TRAIN noise using DPSS Tapers.

OBJ	SNR	THRESHOLDING METHODS -TRAIN NOISE (WAVELET FAMILY-db8, q0=5, L=5, DPSS TAPERS USED)							
		H-TH	S-TH	MI-TH	MIA-TH	MS-TH	MA-TH	TH-I	TH-II
LLR	0 dB	1.2597	1.3246	1.2186	1.1777	1.2726	1.2763	1.2031	1.2153
	5 dB	1.0514	1.1381	1.0303	0.9944	1.0760	1.0797	1.0106	1.0322
	10 dB	0.9516	1.0220	0.8861	0.8663	0.9771	0.9824	0.8904	0.8927
	15 dB	0.8164	0.8850	0.7569	0.7433	0.8415	0.8449	0.7608	0.7672
WSS	0 dB	74.4448	68.7524	74.9075	77.9702	73.3071	72.9520	74.9069	74.3529
	5 dB	58.4855	54.2025	60.2990	62.1080	57.5092	57.1946	59.2396	60.1873
	10 dB	48.5619	44.6301	49.3499	50.7183	48.2338	47.9446	49.1556	48.8809
	15 dB	40.7426	37.2468	41.4893	43.4863	40.5588	40.2432	41.3967	41.0057
CEP	0 dB	6.7796	7.0409	6.6382	6.5100	6.8309	6.8454	6.5869	6.6387
	5 dB	6.0833	6.4042	5.9895	5.8697	6.1622	6.1766	5.9435	6.0040
	10 dB	5.5787	5.8284	5.3604	5.2969	5.6762	5.6951	5.3628	5.3766
	15 dB	5.0951	5.3502	4.8873	4.8230	5.1891	5.2006	4.8980	4.9124
SEG-SNR	0 dB	-2.1229	-2.3997	-1.8297	-1.6853	-2.2023	-2.2220	-1.7969	-1.8625
	5 dB	-0.0010	-0.4286	0.1419	0.3250	-0.1489	-0.1734	0.2204	0.0131
	10 dB	1.5791	1.0260	1.7177	1.9889	1.3393	1.3181	1.7690	1.5368
	15 dB	2.8232	2.3645	2.9758	3.1258	2.6553	2.6384	2.9889	2.8408
FWSEG	0 dB	4.7681	4.5941	4.9327	4.9964	4.6699	4.6790	4.9070	4.9431
	5 dB	6.2601	5.9048	6.4194	6.4353	6.1774	6.1586	6.3359	6.3470
	10 dB	7.4129	7.1844	7.4911	7.4923	7.2303	7.2159	7.4238	7.4050
	15 dB	8.5081	8.3798	8.6114	8.5574	8.3627	8.3710	8.5281	8.4773
SNR	0 dB	2.8988	2.4089	3.3171	3.6410	2.6676	2.6312	3.3889	3.2286

IJAREEIE

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

	5 dB	6.0872	5.2540	6.4995	6.9627	5.7312	5.6887	6.6771	6.2693
	10 dB	7.5751	6.6507	8.0715	8.5373	7.0351	7.0033	8.0301	7.6190
	15 dB	8.8973	8.0485	9.3194	9.7309	8.4458	8.4226	9.3933	9.0157

Table.1(b): Comparative Performance of Thresholding Methods in terms of six Objective Quality Measures LLR, WSS, CEP, SEG-SNR, FWSEG and SNR Measures for real world WHITE noise using DPSS Tapers.

. OBJ	SNR	THRESHOLDING METHODS - WHITE NOISE (WAVELET FAMILY-db8, q0=5, L=5, DPSS TAPERS USED)							
		H-TH	**S-TH**	**MI-TH**	**MIA-TH**	**MS-TH**	**MA-TH**	**TH-I**	**TH-II**
LLR	0 dB	1.3215	1.3514	1.3075	1.3071	1.3350	1.3349	1.2968	1.3096
	5 dB	1.1824	1.2424	1.1451	1.1584	1.2016	1.2038	1.1491	1.1473
	10 dB	1.0261	1.0979	0.9834	0.9796	1.0358	1.0381	0.9708	0.9844
	15 dB	0.8913	0.9489	0.8379	0.8282	0.8931	0.8964	0.8137	0.8337
WSS	0 dB	82.2453	78.8176	84.5702	83.7658	81.8042	81.5223	80.9986	84.0150
	5 dB	67.9707	65.7176	68.4187	68.4650	67.7365	67.5810	66.3124	67.8489
	10 dB	56.7508	54.5500	55.9946	55.3774	56.9785	56.7621	53.9362	55.3996
	15 dB	47.0216	45.0716	45.9590	45.0542	47.1603	46.9765	43.8348	45.3885
CEP	0 dB	7.6562	7.7100	7.7115	7.7582	7.6582	7.6584	7.6520	7.7356
	5 dB	6.9865	7.0951	6.9747	7.1047	6.9947	6.9989	6.9664	6.9834
	10 dB	6.2930	6.4363	6.2598	6.3068	6.2698	6.2768	6.1789	6.2612
	15 dB	5.6765	5.7631	5.5720	5.5849	5.6005	5.6069	5.4270	5.5395
SEG-SNR	0 dB	-0.5691	-0.9735	-0.3542	-0.1226	-0.7004	-0.7194	-0.3151	-0.4375
	5 dB	0.7851	0.2021	1.1129	1.3863	0.6388	0.6052	1.1504	1.0105
	10 dB	1.9331	1.3706	2.3191	2.7837	1.7372	1.7105	2.4355	2.1785
	15 dB	3.0410	2.5342	3.4358	3.7442	2.8907	2.8692	3.5037	3.3351
FWSEG	0 dB	4.7417	4.5485	4.9183	4.9090	4.6779	4.6630	4.8936	4.9284
	5 dB	5.9299	5.7237	6.1665	6.1684	5.8602	5.8601	6.1137	6.1926
	10 dB	7.0995	6.7891	7.4090	7.4654	6.9951	6.9995	7.4242	7.3856
	15 dB	8.2902	8.0593	8.6034	8.6135	8.1921	8.2083	8.5785	8.6077
SNR	0 dB	5.1161	4.2680	5.7254	6.1005	4.7771	4.7350	5.6459	5.5566

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

	5 dB	6.5498	5.4621	7.3090	7.8187	6.1961	6.1391	7.3383	7.1572
	10 dB	7.5036	6.4787	8.3998	9.2976	6.9900	6.9540	8.5653	8.0834
	15 dB	8.4654	7.6677	9.4095	10.0854	8.0329	8.0183	9.5767	9.1927

VII.SUMMARY AND CONCLUSIONS

A new modified Arctan function based Wavelet thresholding speech enhancement method was proposed in this paper. This is based on wavelet thresholding the multitaper spectrum of speech. Wavelet denoising techniques were used in this paper not to remove the noise from the signal, but rather to produce better, lower-variance, spectral estimates and better *a priori SNR* estimates. Informal listening tests revealed that the enhanced speech had no musical noise. The six objective quality measures showed that the proposed method had superior speech quality when compared with the other Threshold methods presented in the work for comparison. The results confirmed the improvement in performance and achievements of our proposed Modified Arctan function based Wavelet thresholding the multitaper Spectra for Speech Enhancement to achieve better results.

REFERENCES

[1]. J Lim & A Oppenheim, Enhancement and bandwidth compression of noisy speech, *Proc IEEE*, vol 67, pp 1586-1604, Dec 1979.
[2]. P. C. Loizou, Speech Enhancement: Theory and Practice, 1st ed. CRC, 2007.
[3] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113–120, 1979.
[4] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 208–211, 1979.
[5] Y. Hu and P. C. Loizou, “Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum”, *IEEE Trans. Speech Audio Processing*, Vol. 12, No. 1, pp.59-67, January 2004.
[6] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, “Speech enhancement based on audible noise suppression,” *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 479–514, Nov. 1997.
[7] N. Virag, “Single channel speech enhancement based on masking properties of the human auditory system,” *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 126–137, Mar. 1999.
[8] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109–1121, 1984.
[9] O. Cappé, “Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor,” *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 345–349, 1994.
[10] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept. 1982.
[11] M. Bahoura and J. Rouat, “Wavelet speech enhancement based on the teager energy operator,” *IEEE Signal Processing Lett.*, vol. 8, no. 1, pp. 10–12, Jan. 2001.
[12]. V.Harika, A.S.R Reddy, S.China Venkateswarlu, “Telugu Speech Enhancement In Terms Of Objective Quality Measures Using Discrete Wavelet Transform With Hybrid Thresholding”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 8, August 2014. Pp.11234-11247 ISSN(Print) : 2320-3765, ISSN(Online):2278-8875
[13]. Y. Ghanbari , M. Karami, ” A Modified Speech Enhancement System Based on the Thresholding of the Wavelet Packets”, 13th ICEE-2005, Vol. 3, Zanjan, Iran, May 10-12, 2005.
[14]. Ting-Hua Yi, Hong-Nan Li and Xiao-Yan Zhao,” Noise Smoothing for Structural Vibration Test Signals Using an Improved Wavelet Thresholding Technique”, Sensors, 12, 2012, pp. 11205-11220, doi:10.3390/s120811205
[15]. David L. Donoho, “De-noising by Soft Thresholding*”*, IEEE Transactions on Information Theory, vol. 41, No. 3, 1995.
[16]. J.I. Agbinya, “Discrete Wavelet Transform Techniques in Speech Processing”, IEEE Tencon Digital Signal Processing Applications Proceedings, IEEE, New York, NY, 1996, pp 514-519.
[17]. Y.Hu & P.Loizou, “Evaluation of objective measures for speech enhancement”.IEEE Trans. Audio speech Lang. process, Vol .16. No.1, pp.229- 238, Jan-2008.
[18]. P. Krishnamoorthy, “An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms”, IETE technical review, vol 28, issue 4, jul-aug 2011, pp 292-301.
[19]. IEEE Subcommittee, IEEE recommended practice for speech quality measurements, IEEE Trans. on Audio Electro acoustics, Vol. 17, Issue 3, pp. 225-246, September 1969.
[20]. The NOIZEUS database: http://www.utdallas.edu/~loizou/speech/ noise
[21]. S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[22]. D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge, MA: Cambridge Univ. Press, 1993.
[23]. K. S. Riedel and A. Sidorenko, “Minimum bias multiple taper spectral estimation,” *IEEE Trans. Signal Processing*, vol. 43, pp. 188–195, 1995.
[24]. P. Moulin, “Wavelet thresholding techniques for power spectrum estimation,” *IEEE Trans. Signal Processing*, vol. 42, pp. 3126–3136, Dec. 1994.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

[25]. A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 3153–3165, 1998.
[26]. A. C. Cristán and A. T.Walden, "Multitaper power spectrum estimation and thresholding Wavelet packets versus wavelets," *IEEE Trans. Signal Processing*, vol. 50, pp. 2976–2986, Dec. 2002.
[27]. G. Wahba, "Automatic smoothing of the log periodogram," *J. Amer. Statist. Assoc.*, vol. 75, pp. 122–132, 1980.
[28]. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[29]. D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.
[30]. D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet thrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.
[31]. I. M. Johnstone and B.W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. R. Statist. Soc. B*, vol. 59, pp. 319–351, 1997.
[32]. S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet presentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, pp. 674–693, July 1989.
[33]. Y. Hu and P. C. Loizou, "Ageneralized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 334–341, July 2003.
[34]. Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, 1995.
[35]. S. China Venkateswarlu, ASR Reddy, K. S. Prasad, "Speech Enhancement using Boll's Spectral Subtraction Method based on Gaussian Window", Global Journal of Researches in Engineering: F, Electrical and Electronics Engineering, Volume 14 Issue 6, Version 1.0, pp.9-20,Year 2014,Type: Double Blind Peer Reviewed International Research Journal, Publisher: Global Journals Inc. (USA), Online ISSN: 2249-4596 & Print ISSN: 0975-5861.
[36]. D. Klatt, "Prediction of perceived phonetic distance from critical band spectra," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,1982, vol. 7, pp. 1278–1281.
[37]. Quackenbush,T.Barnwell,M.Clements,"objective measures of speech quality", Engle wood Cliffs, NJ: Printice Hall, 1988.

Biography

S.China Venkateswarlu born in Mittathmakur, a nearby Village of Gudur, Nellore District, Andhra Pradesh, India in 1969. He obtained his B.Tech (Electronics & Communication Engineering) from Nagarjuna University (Acharya Nagarjuna University), A.P, INDIA. He earned his DPH & S.Tech from the College of Medical Technology, Ongole, A.P, India. M.Tech (Digital Systems & Computer Electronics) from JNTU College of Engineering, Kukatpally, Hyderabad, A.P, INDIA. And Pursuing PhD degree from JNTU College of Engineering, Kakinada, A.P, INDIA. All in Electronics & Communication Engineering. He Worked as Associate Lecturer, Assistant Professor, Associate Professor, Professor, and Head of the Department in different Engineering Colleges of Andhra Pradesh, INDIA. Worked as an Professor in Dept. of ECE of MREC-Hyderabad, TG, INDIA. Presently he is working as an Associate Professor in Department of Electrical & Computing Engineering, from Adama Science & Technology University, Adama, Ethiopia., S.China Venkateswarlu is having more than 16 years of teaching experience and published more than 25 papers in referred National and International Journals. He has presented more than15 papers in National and International Journals. He has reviewed one Book on Digital Signal Processing from M/S Persons Educations. He is a Life member of ISTE, CSI, and IAENG. He is working as a Reviewer in Different International Journals. His areas of interest Digital Signal Processing, Speech Processing, Embedded Systems, Microcontrollers ,Hardware Programming and Cellular & Mobile Communications etc., E-mail IDs: sonagiricv@gmail.com, cvenkateswarlus@gmail.com.

Dr. A. Subbarami Reddy born in Anjimedu, a nearby village of Tirupati, Andhra Pradesh, India in 1958. He obtained his M.Sc Physics (Electronics) from S.V University, Tirupati, India. He earned his AMIE from the Institution of Engineers (India), Kolkata, India, M.Tech from now NIT, Kurukshetra, India and PhD degree from Andhra University, Waltair, India all in Electronics and Communication Engineering. He worked as Laboratory Assistant, Associate Lecturer, Lecturer, Assistant Professor, Associate Professor, Professor, Sr.Professor and Head of the Department and Dean in different Engineering colleges of Andhra Pradesh, India. Presently he is the Principal of Srikalahasteeswara Institute of Technology (SKIT), Srikalahasti, Andhra Pradesh, India. Dr. A. Subbarami Reddy is having more than 30 years of experience and published more than 25 papers in referred International and National Journals. He is a life member of ISTE (India), Fellow of IETE. His areas of interest include Signal Processing and its Sub fields. Email ID: skitprincipal@gmail.com

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 1, January 2016

K.SatyaPrasad[3], B.Tech., ME., Ph.D., L.M.I.S.T.E., MIEEE, F.I.E., F.I.E.T.E. is a Ex-RECTOR & PROFESSOR in ECE, in JNT University, Kakinada, A.P, INDIA, Worked in various Engineering College and have 38 years of Teaching, R&D and Industrial experience. He has published 28 research papers in National and International Journals. He has presented 20 National and 2 International Conference Papers. His research interest includes Signal Processing, Speech Processing, Wireless Communications and Embedded Systems. He has guided 04 PhDs, 15 students pursuing PhD under his guidance. He has guided 15 M.Techs and 13 Projects. He has authored or Co-authored book entitled: Electronic Devices and Circuits. He is the Member Curriculum development: Chairman, Board of Studies, JNTU, Hyderabad and GPR College of Engineering, Kurnool. He officiated as : Head of ECE dept. from January 1997 to September 2003, Vice-Principal, JNTUCEK from October 2003 to May 2005 Principal, JNTUCEK from May 2005 to June 2007. He has completed R & D Project sanctioned by Ministry of Human Resources and Development, Government of India. Email ID: prasad_kodati@yahoo.co.in.