Adaptive Thresholding using Median Base Filtering for Quality Improvement of Signal

Muhammad Kashif, Dr.Sheeraz Ahmed, Fayyaz A.Chaudhary

***Abstract*—**Speech signal segmental framing and scaling factor according to the low distortion of original speech signal is basis for speech recognition process as first step. The next followed step is existing noise reduction in the recognized speech signal for quality improvement. In this work, the noise reduction is done using newly proposed adaptive median based filtering. Comparison of the observations based on adaptive median filtering with Minimum Mean-Square Error Short-time Spectral Amplitude (MMSE-STSA) and Minimum Mean-Square Error (MMSE) based noise reduction reveal a list of worthy to mention relevant observations. The drawn conclusion also accumulates possible contributions by the proposed adaptive median based filtering technique. Lastly is mentioning of Signal-to-noise ratio (SNR) as the primary metric for observations collection for the newly proposed adaptive median based filtering technique analysis.

Keywords*— *MMSE, SNR, STSA, FFT, AWGN, SNR

I. INTRODUCTION

Speech enhancement main purpose is to upgrade the speech signal intelligibility and quality. The intelligibility is the ability of listener to understand the content of speech signal. The intelligibility varies from person to person but the minimum speech reception threshold is fifty (50) percent [1]. Whereas quality is the comparison of system output speech signal with the original input signal. Quality may vary from person to person. Some person mark low quality speech as a high quality and other person mark it low quality. So quality is measure of speech signal better effect on listener ear. Both of them are independent attribute of speech signal. Mostly both are inversely related with each other [2, 3].

Speech signal comprises of both useful and unwanted signals. A common definition, the unwanted signal part of speech signal is called noise. In presence of background noise, the speech signal becomes degraded by noise. An important worthy to note relevant discrepancy is domination of speech signal by the noise signal and this occurs with occurrence of negative SNR which results in zero intelligibility. However, the noise can be reduced easily if attributes of noise or speech signal are known. Achieving limited speech signal distortion and noise quantity, different speech enhancement techniques are used. The list includes Speech enhancement using MMSE filter in wavelet domain [4], Speech enhancement using Empirical Mode Decomposition method [5], adaptive median based filtering for quality improvement [6].Enhancement technique works when the attributes of noise are known. List of different types of noise includes pink noise, babble noise, grey noise and additive noise [7]. Additive noise is the background noise which has a resemblance with speech signal. The best example of additive noise is additive white Gaussian noise. Additive white Gaussian (AWGN) noise is used as a basis noise for adaptive thresholding using median base filtering for quality improvement of signals[8]. Improve signal has many application such as hearing impaired devices. Real time analysis of speech enhancement is too much expensive and not in range. Alternatively the mathematical analysis has solved the purpose with significant quality which can be seen in subsequent sections of the document.

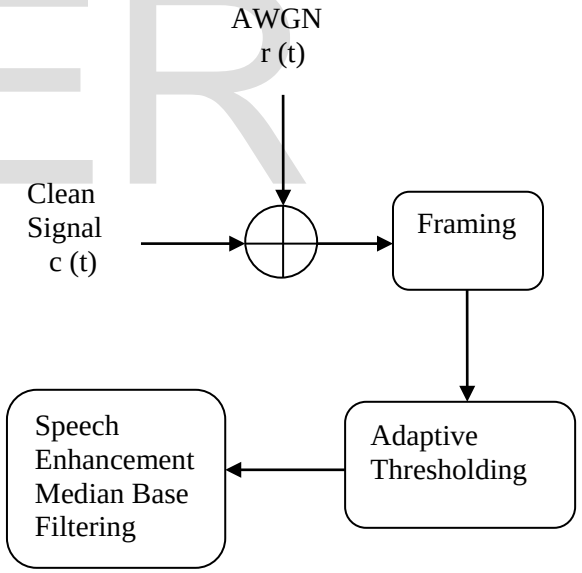

Figure 1.1 Block diagram of binary mask method

The speech enhancement can be done using different proposed method. These methods comprises of parametric and non-parametric method[9]. The parametric model method doesn't need the posterior information[10]. The priori model information helps in the process of speech enhancement with reduced noise and better quality. For single channel the non-parametric method use the silent frame as frame of reference of speech signal and estimate the noisy observation. However, for multiple channels one source is assumed to be reference source for noise estimation. Single channel is more effective because of its

clear implementation and least cost[11]. Also, it is important to note that non-parametric approach is flexible and therefore allow changes in the used parameters over a cartage specified range therefore not restricted to fixed number of parameters as the case in parametric approach.
Here in this case, adaptive thresholding using median base filtering algorithm is based on non-parametric and single channel speech enhancement. The algorithm needs the noise observation of silent frame of speech signal which is achieved using single microphone. In the following text the survey of few speech enhancement algorithms provides the basis for proposed algorithm.

II. Literature Review of Speech Enhancement Method

Speech enhancement comprises of three different algorithm categories[12]. The algorithms categorization depends upon pros and corns in their operation. Four (04) different speech enhancement categorizations are binary mask, spectral subtraction, subspace algorithm, statistical based algorithms.

A. Spectral Subtraction method

Spectral Subtraction needs the continuous output for reduction of noise. The spectral subtraction method for the first time induces by the Weiss using in the correlation domain. Later Boll uses it for in the Fourier transform domain. The spectral subtraction method is based on subtraction of noise from the corrupted signal[13]. During speech enhancement using spectral subtraction method first determine the noise. The noise can be determined from the pause region of the speech signal. And subtraction of the highlighted noise is carried out from the original signal. However, this algorithm can distort the useful information from the speech signal.

B. Binary mask method

The binary mask method use the binary value for noise reduction. In binary masking method original signal is added with selected frequencies of corrupted signal. After addition of noise the time domain signal is converted into the time-frequency domain. The frequency domain analysis of signal provides the frequency information of speech signal. The frequency domain not provides the change of energy of speech signal at specific point. So it needs the use of time domain. As the speech signal is non-stationary so it need the use of both frequency and time domain. Classical method of frequency domain is Fourier analysis but for the binary masking method represented by Short time Fourier transform[14]. In time and frequency domain the binary masking & signal-to-noise ratio thresholding criterion is created. The difference between the target energy signal and energy of mask signal is greater than local threshold criterion than assign value 1.If it is not greater than the local threshold criterion assign zero.

$$BM(t,f) = \begin{cases} 1 & if TE(t,f) - m(t,f) > LTC \\ 0 & otherwise \end{cases}$$

The binary masking value is multiplied with the magnitude of FFT. The output from the FFT is converted into inverse FFT and the result is weighted by using overlap add method and apply windowing and get output frames in row frame matrix. The block diagram is shown in figure 1.1

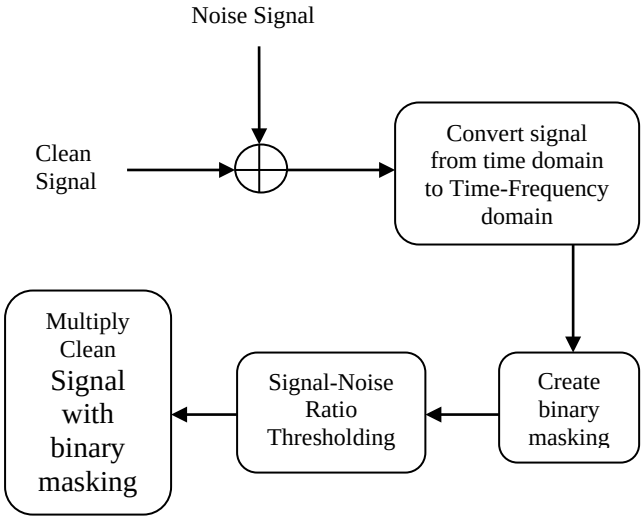

Figure 1.2 Block diagram of binary mask method

C. Subspace method

In speech enhancement using subspace method noisy speech signal is the sum of speech signal subspace and the additive noise subspace[15]. As a first step speech signal is converted into subspace covariance matrix for the noisy speech signal. The noisy speech signal is the sum of the covariance matrix of speech signal and covariance matrix of additive noise signal. Then estimation of the speech signal subspace covariance matrix is next followed step. The Eigen value decomposition method is used for finding the diagonal element greater than zero of covariance matrix of signal subspace and noisy subspace. Then get the signal subspace and noisy subspace spanning the Eigen vector against the Eigen values. The diagonal element of covariance matrix of speech signal provides the dimension of covariance matrix speech signal. After this conversion of the product of Eigen value and Eigen vector into frequency domain occurs. In frequency domain the power spectral density of Eigen Product divided by the number of samples of the frames is found. Calculation of the tonal and non-tonal component using the thresholding masking is the next in-line step of the process. For minimizing the noise in the enhance speech signal the noise frame with the maximum value of the noise is normalized in the frame. Now after normalization is the time for back into Eigen domain conversion from converted frequency domain. Calculation of the gain and the product of noise frames and filter provide the estimation of speech enhancement.

D. Statistical based Method

Initially for statistical base speech enhancement method the Ephraim used the MMSE STSA method[16]. MMSE STSA is primarily used for the noise estimation. The noisy speech signal is divided into magnitude and phase part. The estimation of speech signal depends on the product of response of the MMSE STSA estimator and frame index with sampling frequency. The gain or response of MMSE depends on the priori SNR. The priori SNR depend on the

ratio of speech variance to variance of noise known as posterior SNR or instantaneous SNR. Since It is here that conversion of non-stationary speech signal into stationary occurs by estimating the one fixed frequency and time required in the stationary signal. Point worthy to note is improved quality signal results by the conversion into stationary signal using the newly proposed clean adaptive thresholding that is based on median based filtering.

For illustration of the newly proposed median based adaptive thresholding, priori discussion regarding the primary MMSE method is necessary. This is presented in the following discussion.

III. Overview of MMSE Method

For better explanation of the proposed median based filtering technique in this paper, the relevant description of the preliminary used MMSE method needs to be considered in context of stationary and non-stationary speech signal.
The conventional Minimum mean square error (MMSE) method which is not as useful for non-stationary noise as it is usually effective for stationary noise. One of the main reasons for the mentioned inflexibility is tracking the noise variance along time dimension and typical focus on statistical characteristics of the noise[16].

A. Addition of Noise

The additive white Gaussian noise is added as a noise with the clean speech signal, mathematically noisy speech signal can expressed as

$$n(t) = c(t) + r(t)$$

B. Windowing of Speech Signal

The noisy speech signal is further divided into frames using hamming windowing.

$$n(t) = \sum_{j=1}^{c} Q_j(t) + l_m(t)$$

C. MMSE Noise Estimation

For different frame of $Q_j(t)$, the window noise estimation can be done using adaptive parameter of H_j .Successful preprocessing lead to adaptive median base filtering. For the nose estimation using different frame **u** at sampling frequency F_s in presence of adaptive parameter H_j is mathematically expressed using gamma function as basis function given below

$$\widetilde{g_j} = \Gamma[Qj(t); Hj]$$

D.MMSE filtering Estimation

Posterior signal-to-noise ratio is the ratio of square of frames of noisy signal and square of noisy signal using sampling frequency at instantaneous frame u.

$$SNR_{post} = \frac{Q^2\ (Fs,u)}{D^2\ (Fs,u)}$$

Priori SNR can be obtain from posterior SNR, while prior SNR is the sum of maximum value between posterior SNR and zero weight with (1-β) and square of estimation noisy spectral proceeding frame (u-1) at sampling frequency F**s** with square of estimation spectral proceeding frame (u-1) at sampling frequency F**s**.

$$SNR_{priori} = \alpha + (1-\alpha)\, max\big(SNR_{post}(fs,u), 0\big)$$

MMSE filter response is the ratio of priori signal-to-noise of frame **u** at sampling frequency **fs** to priori signal-to-noise of frame u at sampling frequency plus one.
Can be expressed mathematically

$$L\ (fs, u) = \frac{SNR_{priori}}{SNR_{post}+1}$$

Estimated noise spectral is the product of MSSE filter system response L (fs, u) and Qj (fs, u).

$$\widetilde{R_j} = L\ (fs, u)\ Qj\ (fs, u)$$

IV. Overview of proposed Method

A. Adapting filtering using median based filtering

The proposed adaptive median based filtering minimizes the effect of original speech signal distortion. And show more smooth effect during reconstruction of estimated speech signal.
The noisy speech signal is divided into three parts. First comprises of start region with fewer information of speech, so that a noise can be easily subtracted. Due to low distortion ratio of original signal, it is scale with 0.4.The difference between the frames of region 1 and median of frames 1 region and overall median product with scaling factor of 0.4 gives the value of noise region of σ1.So for the first region of noise estimation value can be expressed mathematically

$\sigma 1 = 0.4 \times Median^*(Q_1(t) - Median^* Q_1(t))$

And this is same for region 2 but requires more information of speech. To get smoothing effect its scaling is weighted with 0.8.If any low amplitude noise signal present it can be amplified with larger weight.

$\sigma 2 = 0.8 \times Median^*(Q_2(t) - Median^* Q_2(t))$ same for the third region so the estimated noise level is

$\sigma 3 = 0.4 \times Median^*(Q_3(t) - Median^* Q_3(t))$

Take the average of these three regions

$$\sigma = \frac{\sigma 1 + \sigma 2 + \sigma 3}{3}$$

And finally the value of σ is use in the equation of Donoho

$$A = \sqrt{2logL}\ \sigma$$

In above equation L is the length of original speech signal.
The block diagram of proposed method is shown in figure 1.2

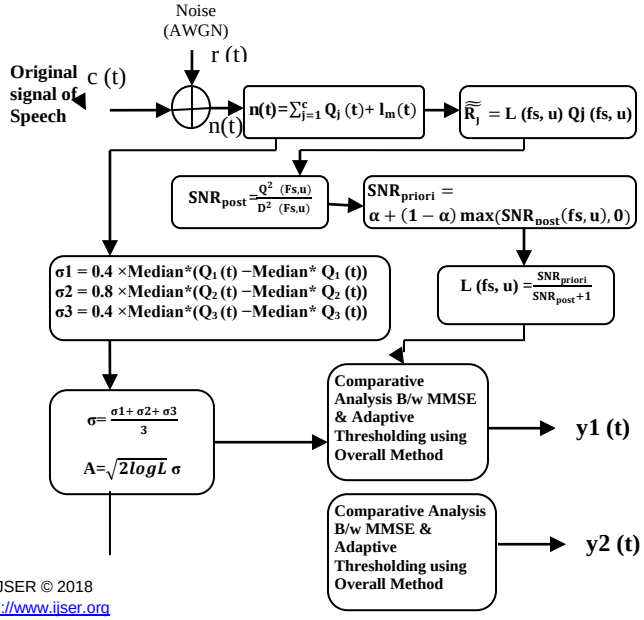

Figure 1.3 Adaptive thresholding using median based filtering block diagram.

V. Experimental Result

Random test are perform for comparative analysis of MMSE and adaptive median based filtering. For initial SNR the output of change in SNR is calculated for the SNR range between -2 dB to -18 dB shown in figure 2.1 and figure 2.2.

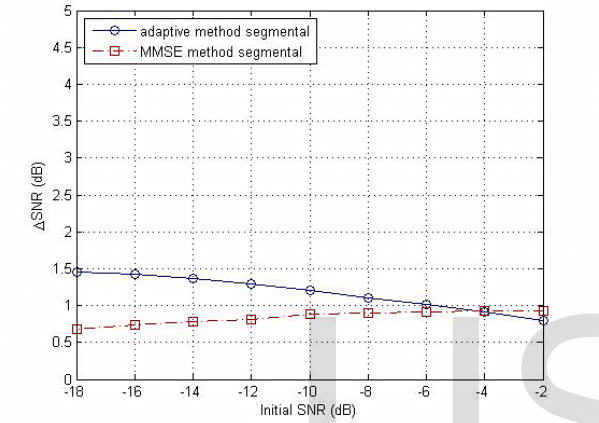

Figure2.1: Segmental result of ΔSNR comparison with MMSE for SNR range between -2dB to -18 dB.

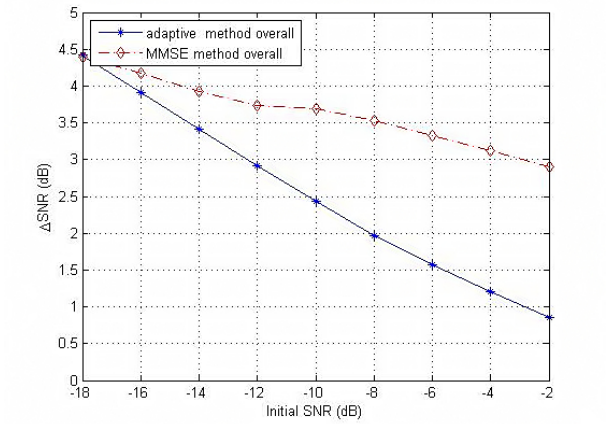

Figure2.2: Overall result of ΔSNR comparison with MMSE for SNR range between -2dB to -18 dB.

The adaptive thresholding using median based filtering show better performance between the ranges of SNR -14 dB -8dB for the input SNR range -2dB to -18dB.As compare to segmental result the proposed method show better performance result for overall analysis. Greater the input range of SNR the better the performance result of proposed method.

Another set of experiment for SNR range 4dB to -16dB.Change in SNR for the input value of initial SNR is calculated shown in figure 3.1 and 3.2.

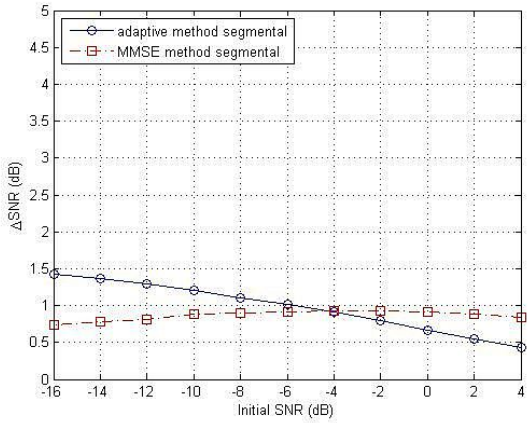

Figure 3.1: Segmental result of ΔSNR comparison with MMSE for SNR range between 4dB to -16dB.

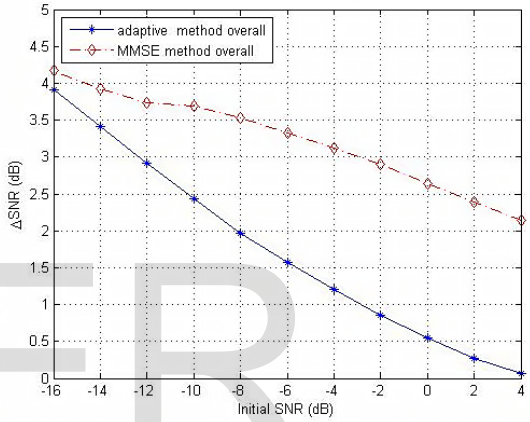

Figure: 3.2 Overall result of ΔSNR comparison with MMSE for SNR range between 4dB to -18 dB.

VI. Conclusion

In terms of overall signal ΔSNR as compared to adaptive thresholding MMSE filtering technique offers improved performance. The overall performance of adaptive thresholding is better in wide range of SNR and show low performance on short range of SNR. On short range of SNR the MMSE performance is better than adaptive thresholding.

References

1. Koning, R., et al., *Perceptual and Model-Based Evaluation of Ideal Time-Frequency Noise Reduction in Hearing-Impaired Listeners.* IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, 2018. **26**(3): p. 687-697.
2. Khaldi, K., et al., *Speech enhancement via EMD.* EURASIP Journal on Advances in Signal Processing, 2008: p. 873204.
3. Rao, C.V.R., M.R. Murthy, and K.S. Rao. *Speech enhancement using perceptual Wiener filter combined with unvoiced speech—A new scheme*. in *Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE*.
4. Kandagatla, R.K. and P. Subbaiah, *Speech*

enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase-uncertainty. Speech Communication, 2018. 96: p. 10-27.
5. Khaldi, K., A.-O. Boudraa, and A. Komaty, *Speech enhancement using empirical mode decomposition and the Teager–Kaiser energy operator*. The Journal of the Acoustical Society of America, 2014. 135(1): p. 451-459.
6. Nabi, W., et al., *A dual-channel noise reduction algorithm based on the coherence function and the bionic wavelet.* Applied Acoustics, 2018. 131: p. 186-191.
7. Heese, F., et al. *Selflearning codebook speech enhancement*. in *Speech Communication; 11. ITG Symposium; Proceedings of*. 2014. VDE.
8. Djaziri-Larbi, S., et al., *Watermark-Driven Acoustic Echo Cancellation.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018. 26(2): p. 367-378.
9. Kuortti, J., J. Malinen, and A. Ojalammi, *Post-processing speech recordings during MRI.* Biomedical Signal Processing and Control, 2018. 39: p. 11-22.
10. Wang, X., et al., *A comparison of recent waveform generation and acoustic modeling methods for neural-network-based speech synthesis.* arXiv preprint arXiv:1804.02549, 2018.
11. Di Liberto, G.M., et al., *Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia.* NeuroImage, 2018.
12. Fu, J., L. Zhang, and Z. Ye, *Supervised monaural speech enhancement using two-level complementary joint sparse representations.* Applied Acoustics, 2018. 132: p. 1-7.
13. Zhang, Z., et al., *Deep learning for environmentally robust speech recognition: An overview of recent developments.* ACM Transactions on Intelligent Systems and Technology (TIST), 2018. 9(5): p. 49.
14. Yilmaz, O. and S. Rickard, *Blind separation of speech mixtures via time-frequency masking.* IEEE Transactions on signal processing, 2004. 52(7): p. 1830-1847.
15. Wiem, B., P. Mowlaee, and B. Aicha, *Unsupervised single channel speech separation based on optimized subspace separation.* Speech Communication, 2018. 96: p. 93-101.
16. Brown, A., S. Garg, and J. Montgomery, *Automatic and Efficient Denoising of Bioacoustics Recordings Using MMSE STSA.* IEEE Access, 2018. 6: p. 5010-5022.

Muhammad Kashif belongs to Peshawar, KPK, Pakistan. He did Bachelors in Electronic Engineering from International Islamic University Islamabad and Masters in Electrical Engineering (Electronic & Communication) from University of Engineering and Technology, Peshawar. He is working as Lab Engineer in Iqra University Islamabad Campus.

Dr. Sheeraz Ahmed belongs to Peshawar, KPK, Pakistan. He did Bachelors in Electrical Engineering from UET Peshawar and Masters in Mathematics from University of Peshawar. He has been teaching for more than twenty years in different Universities of KPK.

His area of expertise is Computer Networks. Under the supervision of Dr. Nadeem Javaid from COMSATS Islamabad, he completed his Ph.D. from CIIT Islamabad, with the thesis entitled “Towards Cooperative Routing in Underwater and Body Area Wireless Sensor Networks.” Recently, he is working as Associate Professor, Dean Engineering Sciences, Gomal University, D.I.Khan, Pakistan. His research work include Renewable & amp; Sustainable Energy, Energy Management, etc.

Fayyaz A.Chaudary was born in Kasur, province Punjab, Pakistan. He received the B.E. and M.E degree in electronic engineering from the International Ialamic University, Islamabad, Pakistan.

In 2009, he joined The University of Lahore, as a Lecturer, and in 2011 became Campus Engineering lab Incharge. Since July 2012, he has been with the Department of Electrical Engineering, SCET UET Taxila, where he was an Assistant Professor, became Director Quality Enhancement Cell in 2017. His current research interests include Hydroelectric Power Plants, Renewable Energy, Energy Management Systems, and Optimization Techniques.

He is pursuing his PhD. Electrical Engineering from Comsats Institute of Information Technology, Wah Cantt.