Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 84 (2016) 22 – 30

Procedia Computer Science

7th International conference on Intelligent Human Computer Interaction, IHCI 2015

Single Channel Speech Enhancement: using Wiener Filtering with Recursive Noise Estimation

Navneet Upadhyay [a*], Rahul Kumar Jaiswal[a]

[a,]*Department of Electronics & Communication Engineering, The LNM Institute of Information Technology, Jaipur - 302 031, India*

Abstract

This paper discusses the problem of single channel speech enhancement in stationary environments, and proposes Wiener filtering with the recursive noise estimation algorithm. The Wiener filter is a linear estimator and minimizes the mean-squared error between the original and enhanced speech. The algorithm is implemented in the frequency domain and depends on the filter transfer function from sample to sample based on the speech signal statistics; the local mean and the local variance. For the noise estimation, the recursive noise estimation approach is used. In this approach, the noise estimation is done by past and present spectral power values, using a smoothing parameter. The value of smoothing parameter is selected in between [0 1]. For the performance evaluation of the proposed speech enhancement algorithm objective evaluations with informal listening tests are conducted for the speech sentences, pronounced by male and female speakers from the NOIZEUS corpus, degraded by White as well as Pink noise types at different SNR levels. For objective measures, signal to noise ratio, segmental signal to noise ratio, and the perceptual evaluation of speech quality are used. The measures prove that the speech enhanced by proposed algorithm is more pleasant to the human ear for both noise conditions in comparison to the conventional speech enhancement method.

Peer-review under responsibility of the Organizing Committee of IHCI 2015

Keywords: speech enhancement, Wiener filter, noise classification, recursive noise estimation approach.

1. Introduction

Speech is one of the most fundamental means of communication between human to human and human to machine in various fields via automatic speech recognition and speaker identification. The present day speech communication systems are severely degraded due to various types of noises which make the listening task difficult for a direct

* Corresponding author. Tel.: +91-141-519-1700; fax: +91-141-518-9214.
E-mail address: nupadhyay@lnmiit.ac.in

Peer-review under responsibility of the Organizing Committee of IHCI 2015
doi:10.1016/j.procs.2016.04.061

listener and cause inaccurate transfer of information. Therefore, the noise suppression is one of the main motives of various research endeavours in the field of speech processing over the last few decades. The researchers attempted to suppress the noise level of degraded speech without distorting the speech signal and also tried to make a speech more pleasant and understandable to the listener. The main purpose of speech enhancement research is to minimize the degree of distortion of the desired speech signal and to improve one or more perceptual aspects of speech, such as the speech quality and/or intelligibility. These two measures are uncorrelated and independent of each other. A speech signal may be of high quality and low intelligibility and vice versa [5-6].

The classification of speech enhancement system is based on the number of microphones of information available for processing into a single channel, dual channel or multi-channel speech enhancement. Although, the performance of multi-channel speech enhancement is better than that of single channel speech enhancement [6], but the single channel speech enhancement is still an important field because of their simple implementation and effectiveness. The single channel is especially useful in mobile communication applications, where only a single microphone is available due to cost and size considerations. The block diagram of single channel speech enhancement system is shown in Fig. 1.

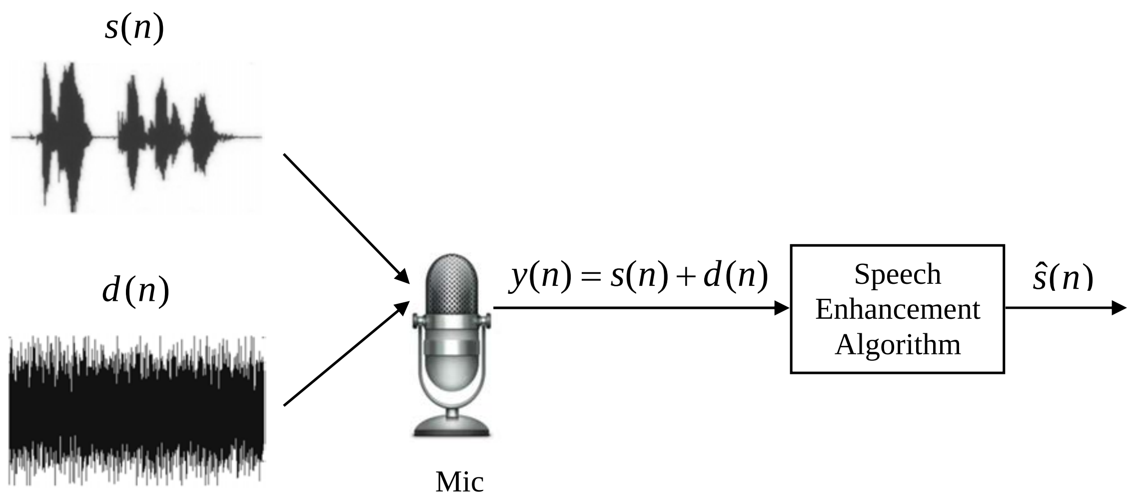

Fig. 1. Block diagram of single channel speech enhancement system.

The spectral subtraction (SS) is one of the most popular and computationally efficient methods for enhancement of single channel speech. The first comprehensive spectral subtraction method was proposed by Boll and is based on the non-parametric approach [1, 5-7]. The SS method exploits the fact that one can obtain an estimate of the clean speech signal spectrum simply by subtracting the noise spectrum from the noisy speech spectrum. But the enhanced signal derived by the SS method is not optimal. Thus, we now turn our attention to Wiener filtering, which is conceptually similar to spectral subtraction but replaces the direct subtraction with an optimal estimate of the clean signal spectrum in a minimum mean square error (MMSE) sense [2-3, 6-8].

In this paper the Wiener filtering with recursive noise estimation algorithm is proposed to enhance the speech degraded by stationary noises. The noise is estimated by first order recursive relation, using a smoothing parameter. The value of smoothing parameter is selected in between [0 1]. The proposed speech enhancement algorithm allows to find the best tradeoff between the amount of noise reduction, the speech distortion and the level of remnant noise in a perceptive view.

The paper is systematized as follows. In Section 2, the review of the spectral subtraction method is described with its connection in filtering domain. In Section 3, the recursive noise estimation approach is described. In Section 4, the

Wiener filtering with recursive noise estimation algorithm is detailed. Finally, an objective evaluation with informal listening tests of the WF-RANS is performed in Section 5.

2. Spectral Subtraction Method

For an implementation of the spectral subtraction method, few assumptions are necessary. Firstly, the speech signal is assumed to be stationary and the noise spectrum does not change significantly in between the update periods; secondly, the speech and noise should be additive and uncorrelated wide-sense stationary (WSS) random stochastic processes with zero-mean [1] and thirdly, the phase of the noisy speech is kept unchanged, since it is assumed that the phase distortion is not perceived by the human ear.

Assume that the noisy speech $\boldsymbol{y}[\boldsymbol{n}]$ can be expressed as $\boldsymbol{y}[\boldsymbol{n}] = \boldsymbol{s}[\boldsymbol{n}] + \boldsymbol{d}[\boldsymbol{n}]$, where $\boldsymbol{s}[\boldsymbol{n}]$ is the clean speech and $\boldsymbol{d}[\boldsymbol{n}]$ is the additive noise. As the enhancement is carried out according to the frame, the above model can be expressed as

$$y(n,k) = s(n,k) + d(n,k), \qquad n = 0,1,2, \ldots, (N-1); \quad k = 1,2, \ldots, N \tag{1}$$

Here n is the discrete time index, k is the frame number and N is the length of the frame.

$$Y(\omega,k) = S(\omega,k) + D(\omega,k) \tag{2}$$

Here ω is the discrete angular frequency index of the frames.

$$Y(\omega,k) = \textstyle\sum_{n=-\infty}^{\infty} y(n)w(k-n)e^{-j\omega n} \tag{3}$$

where $w[n]$ is analysis window, which is time-reversed and shifted by k samples. Multiplying both sides of (2) by their complex conjugates, we get

$$|Y(\omega,k)|^2 = |S(\omega,k)|^2 + |D(\omega,k)|^2 + 2|S(\omega,k)D(\omega,k)| \tag{4}$$

$|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}$ is the short-time power spectrum of speech. The $|\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}$ and $|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|$ can't be obtained directly and are approximate as

$$\boldsymbol{E}\{|\boldsymbol{Y}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} = \boldsymbol{E}\{|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} + \boldsymbol{E}\{|\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} + \boldsymbol{2E}\{|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|\} \tag{5}$$

where $\boldsymbol{E}\{.\}$ represents the expectation operator. As the additive noise is assumed to be zero-mean and uncorrelated, the $|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|$ reduces to zero

$$\boldsymbol{E}\{|\boldsymbol{Y}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} = \boldsymbol{E}\{|\boldsymbol{S}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} + \boldsymbol{E}\{|\boldsymbol{D}(\boldsymbol{\omega},\boldsymbol{k})|^{\boldsymbol{2}}\} \tag{6}$$

Normally, $E\{|D(\omega,k)|^2\}$ is estimated during non-speech activity periods and is denoted by $\hat{P}_d(\omega,k)$. Therefore the estimate of the clean speech power spectrum

$$\hat{P}_s(\omega,k) = P_y(\omega,k) - \hat{P}_d(\omega,k) \tag{7}$$

Here $\hat{P}_s(\omega,k)$ is called an enhanced speech power spectrum, $P_y(\omega,k)$ is the noisy speech power spectrum and $\hat{P}_d(\omega,k)$ is the noise power spectrum which is taken from speaker silence frames [5]. The flow diagram of the spectral subtraction method is shown in Fig. 2.

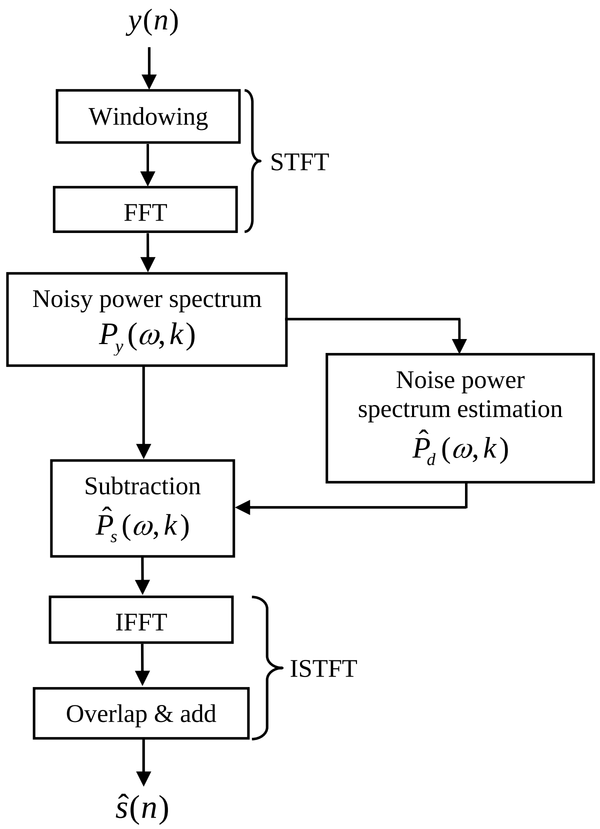

Fig. 2. Flow diagram of spectral subtraction method.

The major drawback of the spectral subtraction method is that the enhanced speech is accompanied by an annoying perceptible tonal characteristic and affects the human listening, known as musical noise [1, 5-7]. This noise is sometimes more disturbing not only for the human ear, but also for speaker recognition systems. Several variations of spectral subtraction method have been proposed to overcome the problem of musical noise [1, 5-7].

The spectral subtraction (7) can be written as

$$\begin{aligned}\hat{P}_s(\omega,k) &= P_y(\omega,k)\left[1-\frac{\hat{P}_d(\omega,k)}{P_y(\omega,k)}\right]\\ &= P_y(\omega,k)H(\omega)\end{aligned} \tag{8}$$

Here, $H(\omega)$ is the filter gain, which is real.

3. Recursive Averaging Noise Estimation Approach

The noise estimation is the most critical part of the frequency domain enhancement algorithm because the quality of the enhanced speech depends on the accurate noise power spectrum estimation. If the noise estimate is too low, annoying musical noise will be audible, and if the noise estimate is too high, speech will be distorted, possibly resulting in intelligibility loss [4].

In our algorithm, the noise estimation is done by averaging past spectral power values, using a smoothing parameter. The noise power is estimated by the first order recursive relation as

$$\hat{P}_d(\omega,k) = \alpha\,\hat{P}_d(\omega,k-1) + (1-\alpha)P_y(\omega,k) \tag{9}$$

where α is a smoothing parameter whose value is selected in between [0 1] and k is the current frame index, ω is the frequency bin index, $P_y(\omega,k)$ is the short-time power spectrum of noisy speech, $\hat{P}_d(\omega,k)$ is the noise power spectrum estimate in ω^{th} frequency bin of current frame and $\hat{P}_d(\omega,k-1)$ is the past noise power spectrum estimate.

4. Wiener Filtering with Recursive Noise Estimation Algorithm

The Wiener filter gives the MMSE estimate of the short-time Fourier transform (STFT) whereas the spectral subtraction obtains the MMSE estimate of the short-time spectral magnitude without changing the phase [2-3, 6-8]. Thus, the condition for minimum mean square error is,

$$H(\omega) = \frac{P_{ss}(\omega)}{P_{yy}(\omega)} = \frac{P_{ss}(\omega)}{P_{ss}(\omega)+P_{dd}(\omega)} \tag{10}$$

Here $\boldsymbol{P_{ss}(\omega)}$ and $\boldsymbol{P_{dd}(\omega)}$ are the signal and noise power spectrum, respectively.

$$H_{Wiener}(\omega) = \frac{P_{ss}(\omega)}{P_{yy}(\omega)} = \frac{P_{yy}(\omega)-P_{dd}(\omega)}{P_{yy}(\omega)} \tag{11}$$

The enhanced signal is estimated as,

$$\boldsymbol{\hat{P}_s(\omega,k) = H_{Wiener}(\omega)P_y(\omega,k)} \tag{12}$$

$$|\hat{s}[n]| = IFFT[\sqrt{\hat{P}_s(\omega,k)}] \tag{13}$$

The estimate of the clean speech spectral magnitude is recombined with the noisy phase giving an estimate of the enhanced speech signal as,

$$\boldsymbol{\hat{s}[n] = |\hat{s}[n]|\angle y[n]} \tag{14}$$

The block diagram of the Wiener filtering with recursive noise estimation algorithm (WF-RANS) is shown in Fig. 3.

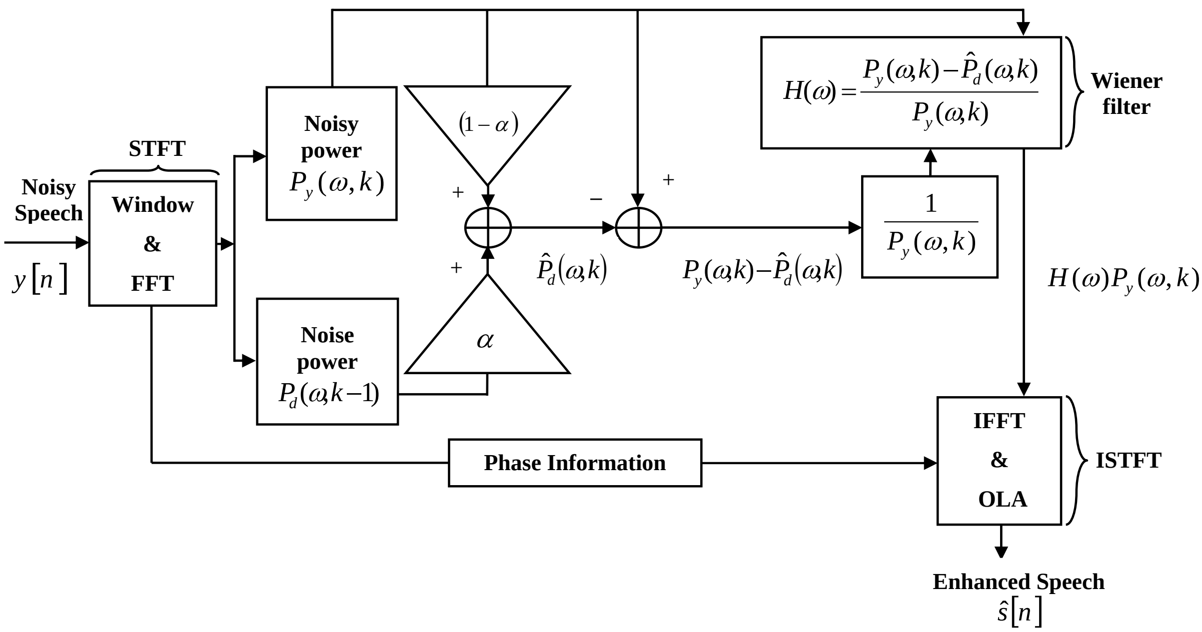

Fig. 3. Block diagram of Wiener filtering with the recursive noise estimation algorithm.

5. Simulation and Results Analysis:

This section presents the simulation results and performance evaluation of the proposed speech enhancement algorithm, Wiener filtering with recursive noise estimation, and its comparison with the spectral subtraction method. For simulations, we have employed MATLab software as the simulation platform. For the experimental purpose, the clean speech samples have been taken from NOIZEUS corpus [9], which is a publicly available speech database and is usually used for benchmark experiments. NOIZEUS comprises of 30 phonetically balanced IEEE sentences produced by six speakers. Also, the speech database is sampled at 8 kHz and quantized linearly using 16 bits resolution. A total of four different utterances (two male speakers and two female speakers), has been used in our evaluation.

The noise signals have different time-frequency distributions and therefore present a different impact on speech. Thus, we have taken a computer generated white Gaussian noise (AWGN) and Pink noise for the evaluation of our speech enhancement algorithm [9]. For AWGN case, we have taken noisy samples from the NOIZEUS database while for the pink noise case, first we have down sampled it to 8 kHz and then have mixed with four speech sentences at different signal to noise ratio (SNR) levels, to prepare the speech database.

In our experiments, the noise samples used are of zero-mean and the energy of the noisy speech samples are normalized to unity. The frame size is chosen to be 200 samples (25 ms—a frame time), with 50% overlapping. The sinusoidal Hamming window with size 200 samples is applied to each frame, individually. The windowed speech frame is then analysed using Fast Fourier Transform (FFT) with length 256 samples. The noise is estimated from the noisy speech, using the first order recursive equation (9).

To test the performance of our speech enhancement algorithm, the objective quality measurements like SNR, segmental SNR (SNRseg.), perceptual evaluation of speech quality (PESQ), and for subjective quality tests, informal listening tests have been used [10-12].

Table 1. SNRseg. and PESQ score of Wiener filtering with recursive noise estimation for AWGN noise at different SNR levels.

SNRseg. of Wiener filtering with recursive noise estimation for AWGN noise at different SNR levels											PESQ score at $\alpha = 0.8$ and $\alpha = 0.9$	
SNR (dB)		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.9$
0	**M1**	6.1467	6.0659	6.0133	6.0399	5.9901	6.0175	6.0454	**6.0616**	6.1187	**1.6303**	1.1772
5		6.3677	6.2702	6.2695	6.2913	6.3015	6.3206	6.3580	**6.4312**	6.5102	**2.0182**	1.6995
10		6.5676	6.5376	6.5385	6.5256	6.5806	6.5890	6.6768	**6.7096**	6.8621	**2.2176**	2.0143
15		6.8252	6.7900	6.7987	6.7972	6.8166	6.8741	6.9444	**6.9868**	7.1393	**2.1856**	2.0428
0	**M2**	6.2204	6.1461	6.0805	6.0686	6.0728	6.0423	6.0459	**6.0922**	6.0729	**1.1713**	1.0589
5		6.3976	6.3580	6.3465	6.3312	6.3470	6.3469	6.3834	**6.4227**	6.4842	**1.6756**	1.6657
10		6.6059	6.5684	6.5739	6.5787	6.5155	6.6381	6.6834	**6.7625**	6.8366	**2.1777**	2.1624
15		6.8084	6.7695	6.7817	6.8107	6.8579	6.9055	6.9370	**6.9716**	7.0725	**2.3571**	2.3358
0	**F1**	5.9614	5.7786	5.6670	5.5736	5.4823	5.4748	5.4169	**5.4051**	5.4607	**0.9820**	1.0797
5		6.2681	6.1205	6.0240	5.9785	5.9385	5.9351	5.9432	**5.9614**	6.0864	**1.9816**	1.7841
10		6.5369	6.4735	6.4010	6.4322	6.3949	6.3962	6.4655	**6.4918**	6.5995	**2.3012**	2.3000
15		6.7649	6.7180	6.7246	6.7218	6.7190	6.7818	6.7973	**6.8897**	6.9835	**2.3712**	2.3602
0	**F2**	6.4262	6.3111	6.2797	6.2683	6.2425	6.2454	6.2461	**6.2627**	6.2864	**1.2494**	1.1830
5		6.6165	6.5597	6.5419	6.5284	6.5407	6.5734	6.5902	**6.6821**	6.7447	**1.8662**	1.8383
10		6.8300	6.8104	6.8302	6.7961	6.8220	6.8639	6.9143	**6.9915**	7.0786	**2.2525**	2.2472
15		6.9911	7.0216	7.0443	7.0635	7.0785	7.1195	7.1900	**7.2309**	7.3318	**2.3947**	2.1773

Table 2. SNRseg. and PESQ score of Wiener filtering with recursive noise estimation for pink noise at different SNR levels.

SNRseg. of Wiener filtering with recursive noise estimation for pink noise at different SNR levels											PESQ score at $\alpha = 0.8$ and $\alpha = 0.9$	
SNR (dB)		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.9$
0	**M1**	7.1804	7.6478	8.1224	8.6437	9.2580	10.0242	10.9384	**12.1556**	14.1045	**2.2797**	2.2561
5		7.5343	8.0645	8.6094	9.2089	9.9063	10.7526	11.8025	**13.1289**	15.1599	**2.2900**	2.2246
10		7.8623	8.4459	9.0351	9.6901	10.4369	11.3619	12.4804	**13.8679**	15.9073	**2.3506**	2.3502
15		8.1451	8.7907	9.4050	10.1007	10.8864	11.8497	12.9939	**14.4112**	16.4543	**2.3128**	2.3722
0	**M2**	7.4840	8.0469	8.6217	9.2468	9.9331	10.7268	11.7171	**12.9963**	14.4753	**2.3540**	2.3487
5		7.8239	8.4938	9.1513	9.8675	10.6583	11.5395	12.5952	**13.9591**	15.4976	**2.4514**	2.4250
10		8.1056	8.8379	9.5594	10.3449	11.1972	12.1281	13.2258	**14.6550**	16.2260	**2.5167**	2.4523
15		8.3597	9.1274	9.8854	10.6985	11.5778	12.5605	13.6931	**15.1609**	16.7699	**2.4551**	2.4262
0	**F1**	7.1277	7.5197	7.9687	8.5086	9.1902	10.0361	11.0740	**12.2320**	13.5888	**2.4532**	2.5074
5		7.6803	8.2478	8.8196	9.4581	10.2177	11.1460	12.2515	**13.3636**	14.6915	**2.5948**	2.5671
10		8.0851	8.7913	9.4689	10.1752	10.9703	11.9247	13.0274	**14.0961**	15.4108	**2.5813**	2.5442
15		8.3932	9.1846	9.9260	10.6615	11.4682	12.4157	13.5044	**14.5681**	15.8823	**2.5331**	2.5087
0	**F2**	7.6561	8.1930	8.7320	9.2916	9.9363	10.7461	11.7180	**12.8056**	14.0104	**2.4605**	2.4499
5		8.0090	8.6549	9.2737	9.9188	10.6028	11.4735	12.5068	**13.6579**	14.9412	**2.4708**	2.4466
10		8.2927	9.0171	9.7005	10.4007	11.1334	12.0158	13.0696	**14.2717**	15.6377	**2.3723**	2.3414
15		8.5195	9.3201	10.055	10.7813	11.5283	12.4203	13.4850	**14.7086**	16.1560	**2.3317**	2.2999

From the extensive study, it is observed that for every case of input SNR (discussed in this paper), the value of smoothing parameter α increases and the value of SNRseg. is better for both AWGN and Pink noise. This is also supported by PESQ score, which is the objective measure of subjective speech quality.

As described in (9) that the noise estimation approach in the current frame is heavily dependent on noise in previous frame as well as lightly dependent on noisy speech in current frame. Therefore, from the different values of α, shown in Table 1 and Table 2 , $\alpha = 0.8$ is the suitable value for our algorithm.

Table 3. Output SNR, Output SNRseg. and Perceptual evaluation of speech quality (PESQ) measure results of enhanced speech signals at (0, 5, 10, 15) dB SNRs. English sentence "*The line where the edges join was clean*", produced by M1 speaker, "*The sky that morning was clear and bright blue*", produced by M2 speaker, "*The set of china hit the floor with a crash*", produced by F1 speaker and "*She has a smart way of wearing clothes*", produced by F2 speaker is used as original signal.

Noise Type		Speech Enhancement Algorithm	SNR (dB)				SNRseg. (dB)				PESQ score			
			0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
M1	AWGN	SS	1.415	2.177	2.942	4.214	6.066	6.440	6.724	6.986	1.0502	1.6001	1.8948	1.9426
		WF-RANS	1.441	2.210	2.944	4.245	6.079	6.444	6.765	7.025	1.6404	2.0182	2.2176	2.1856
	PINK	WF-RANS	4.248	4.442	4.482	4.407	12.555	14.128	14.867	14.411	2.2197	2.2900	2.4506	2.4128
M2	AWGN	SS	2.105	4.542	4.504	5.046	6.072	6.449	6.776	7.001	1.0011	1.6595	2.1124	2.1925
		WF-RANS	2.056	4.462	4.460	5.042	6.067	6.422	6.774	6.991	0.9414	1.6756	2.1177	2.2571
	PINK	WF-RANS	5.174	5.296	5.446	5.449	12.996	14.956	14.655	15.160	2.4540	2.4514	2.5167	2.4551
F1	AWGN	SS	2.289	4.844	5.011	5.564	5.426	5.981	6.497	6.849	0.9950	1.6768	2.1407	2.4589
		WF-RANS	2.405	4.900	4.964	5.641	5.448	6.000	6.557	6.877	0.9820	1.6816	2.2012	2.4712
	PINK	WF-RANS	5.750	5.885	5.940	5.944	12.242	14.464	14.096	14.568	2.4542	2.5948	2.5814	2.5441
F2	AWGN	SS	1.548	2.709	4.418	4.900	6.250	6.654	6.977	7.244	1.1425	1.8454	2.0407	2.2450
		WF-RANS	1.776	2.795	4.416	4.944	6.264	6.674	6.991	7.245	1.1494	1.8062	2.1425	2.4947
	PINK	WF-RANS	4.978	4.074	4.107	4.115	12.805	14.657	14.271	14.708	2.4605	2.4708	2.4724	2.4417

The objective results SNR, SNRseg. and PESQ score, shown in Table 3, is for stationary noises. The SNRseg. is higher than the overall SNR for the both AWGN and Pink noise at different input SNR levels. The overall SNR and SNRseg. is better for our proposed algorithm, WF-RANS, in comparison to SS method. Also, the PESQ, which is an objective evaluation of subjective speech quality, score is higher in the case of Pink noise as compared to AWGN noise. It is also concluded from the PESQ score that WF-RANS algorithm outperforms depending on the input SNR and the improvement in the enhancement of degraded speech is better for female speaker than male speaker and even more better in Pink noise case.

From the SNR point of view, for both AWGN and Pink noise cases, the speech enhancement for female speaker is better than male speaker. Also, from SNRseg. point of view, for AWGN noise case, the speech enhancement for female speaker is better than male speaker. But for Pink noise case, it is almost same to male speaker.

Usually, a speech enhancement system produces two main undesirable effects: musical noise and speech distortion. However, these effects are difficult to quantify with the help of traditional objective measures. Therefore, an informal listening test is conducted in this study for observing both the remnant noise and the speech distortion. It can be observed that the remnant structure of musical noise is reduced more by WF-RANS, as compared to SS method. Therefore, speech enhanced by the WF-RANS is more pleasant and the musical noise, if any, has a "*perceptual white quality*" while distortion remains acceptable. This is confirmed by the values of SNR, SNRseg., and PESQ score and also validated by informal listening tests.

6. Conclusion:

In this paper, the Wiener filtering with recursive noise estimation algorithm was implemented to deal the enhancement of speech degraded by additive noise. For the performance evaluation of the WF-RANS, the objective and subjection measures were conducted at different SNR levels for AWGN and pink noise types and compared with the spectral subtraction method. Results show that the quality of speech enhanced by WF-RANS is good and the musical noise is less structured than the spectral subtraction method, while the distortion of speech remains acceptable.

References

[1]. Boll Steven F., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113-120, 1979.

[2]. Xia Bingyin and Changchun Bao, "Wiener filtering based speech enhancement with Weighted Denoising Auto-encoder and noise classification," *Speech Communication*, vol. 60, pp. 13-29, 2014.

[3]. El-Fattah and Marwa A. Abd, "Speech enhancement with an adaptive Wiener filter," *International Journal of Speech Technology*, vol. 17, no. 1, pp. 53-64, 2014.

[4]. Martin Rainer, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, 2001.

[5]. Navneet Upadhyay and Abhijit Karmakar, "The spectral subtractive-type algorithms for enhancing speech in noisy environments," *IEEE International Conference* on *Recent Advances in Information Technology*, ISM Dhanbad, India, March 2012, pp. 841-847.

[6]. Philipos C. Loizou, Speech Enhancement: Theory and Practice, 2nd edition, CRC Press, 2013.

[7]. Saeed V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, 3rd ed., Wiley, February 2006.

[8]. S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Upper Saddle River, 2003.

[9]. A Noisy Speech Corpus for Assessment of Speech Enhancement Algorithms. *http://ecs.utdallas.edu/loizou/speech/noizeus/* and *http://www.speech.cs.cmu.edu/*.

[10]. Yi Hu and Loizou P. C., "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no.1, pp. 229-238, Jan.2008.

[11]. Perceptual evaluation of speech quality (PESQ), An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation P.862.1, 2003.

[12]. Hu. Yi and Philipos C. Loizou, "Subjective comparison of speech enhancement algorithms," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 14-19, 2006.