Noise Reduction and Speech Enhancement Using Wiener Filter

Hilal H. Nuha
School of Computing,
Telkom University
Bandung, Indonesia
hilalnuha@telkomuniversity.ac.id

Ahmad Abo Absa
University of Palestine,
Palestine
a.absa@up.edu.ps

2022 International Conference on Data Science and Its Applications (ICoDSA) | 978-1-6654-8665-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICoDSA55874.2022.9862912

Abstract*— Digital data transmission rate may reach over 2.5 Tb/s using the orthogonal frequency division multiplexing (OFDM). Digital speech enhancement is crucial during the pandemic era. This is due to most of information and communication is performed online. However, not all people have private room form digital communication. Therefore, background noise from the indoor condition may distort the speech during the recording. Speech denoising has many benefits for instance in voice communication or voice recognition where fast denoising process are needed. This paper evaluates the use of Wiener Filter for noise reduction. Enhancement of distorted speech by additive noise with only single observation has been done and still a challenging problem. We add the noise to the sample clean speech to obtain noisy speech. We generate noise level for SNR 0 up to 0.5dB with increment 0.01dB. We choose low SNR to represent high additive noise. We further apply Wiener Noise Reduction to the noisy speech to obtain filtered noisy speech. Finally, we compare the Mean Square Error (MSE) of filtered speech and the original speech for every noise level. The results show that the noise has been decreased. The non-speech parts now appear better since the noisy part have been suppressed. Our experiment shows that the proposed technique successfully improves the speech in noisy environment up to order of 10^{-4}.

Keywords*— *wiener filter, noise reduction, speech enhancement

I. INTRODUCTION

Digital data transmission rate may reach over 2.5 Tb/s using the orthogonal frequency division multiplexing (OFDM) [1]. Despite the use high transmission rate technology, the needs for digital speech transmission is still very high. Due to high demand, high bandwidth capacity may still suffer from loss that leads to signal distortion. Therefore, the needs of speech signal enhancement is still highly required.

Enhancement of distorted speech by additive noise with only single observation has been done and still a challenging problem. Speech denoising has many benefits for instance in voice communication or voice recognition where fast denoising process are needed [2]. This problem can be viewed as reducing noise from a single voice receiver where noise suppression depends on the frame spectral gain as a function of the a priori Signal to Noise Ratio (SNR) and or the a posteriori SNR as discussed in [3].

Chen et al discuss a new insight on the Wiener Filter can be seen in the paper [4]. The main challenge of using the Wiener Filter for Electro Cardio Graph (ECG) denoising can be seen in the paper [5]. Therefore, the Wiener Filter technique for EEG artifact removal [6] has attracted the attention of researchers. Research related to the Wiener Filter for SAR image despeckling [7] can also be seen in the article. The application of the Wiener Filter method for neural network speech enhancement system is also discussed in more depth in the scientific paper [8]. [9]'s paper also uses the Wiener Filter technique for ultrasound image restoration. The Wiener Filter method is also used for removing eye-blink artifacts from EEG data in scientific papers [10]. Studies on the Wiener Filter for the identification of multilinear forms can also be seen in the paper [11]. The study of the implementation of the Wiener Filter robust multi-channel signal processing can be discussed in the paper [12]. These scientific works show that the Wiener Filter method is suitable for GPS position time series denoising and various [13] problems. Nuha and Suwastika used the Fractional Fourier Transform to reduce the noise level of seismic trace [14]. Since speech and sound quality is important for recognition tasks [15][16], obtaining high quality audio signal is crucial. Therefore, this paper proposed the used of multiple step speech enhancements for audio signal.

II. METHODOLOGY

This section consists of two main discussions. First, the Wiener filter process is described. Second, the experimental setup is provided.

A. Wiener Filter

For our experiment, we use the formulation given by [2]. As the inputs, the process requires noisy speech y_N, sampling frequency f_s, and initial silence M_{IS}. Initial silence M_{IS} is the a priori information regarding number of silence in the speech in term of sample. The speech improvement process begins with determining window length W, where W can be obtained from

$$W = [0.02 \times f_s] \quad (1)$$

This paper used window with size of 0.02s, f_s denotes sampling frequency of the speech data, and $[\cdot]$ is rounding operator. In other words, W is the number of sample within 0.02s. then, we generate Hanning window H_W with size W as a column vector [17]. We calculate noise estimate $\hat{d}$ as the following

$$d = \frac{1}{C}\sum_{c=0}^{C} D(c,W) \quad (2)$$

where $D(c,W) = \{\left|\left|\mathcal{F}(s(c)\ s(c+1) \dots s(c+W))\right|\right|^2\}$, $\mathcal{F}$ is Fast Fourier Transform with $2W$ points, $C = M_{IS} - W$, and $s(n) = y_N(n) \times H_W(n)$.

To obtain Two-step Noise Reduction (TSNR) result we start by initializing the loop L times.

$$L = \left\lceil \frac{1 - 2 \times W}{W - k} \right\rceil \quad (3)$$

where $\phi = 0.25$ is shift percentage for overlap-add method and $k = [(1 - \phi) \times W]$. With $\alpha = 0.98, A'_{-1} = O$, $y'_{l-1} = O$, and $l = 0$ to L, the following:

$$y_l = \{y_N(l \times k + 1) \dots y_N(l \times k + W)\} \quad (4)$$

The speech segment is extracted and the fast fourier transform $\mathcal{F}(\cdot)$ is applied.

$$s_l(n) = y_l(n) \times H_W(n) \quad (5)$$

$$S_l = \mathcal{F}(s_l) \quad (6)$$

The signal phase is further extracted to its magnitude and phase parts.

$$\Theta_l = \angle(S_l) \quad (7)$$

$$A_l = |S_l| \quad (8)$$

$$\eta_l^p(n) = \max\left(\frac{|A_l(n)|^2}{d(n)} - 1, 0.1\right) \quad (9)$$

The noisy speech frame duration is set to be equal to L1 Samples. Where the frame is further converted.

$$\eta_l^a(n) = \frac{\alpha\left(A'_{l-1}(n)\right)^2}{d(n)} + (1 - \alpha)\,\eta_l^p(n) \quad (10)$$

$$A''_l(n) = \frac{\eta_l^a(n)}{\eta_l^a(n) + 1} \times A_l(n) \quad (11)$$

$$T_l(n) = \frac{A''^2_l(n)}{d(n)} \quad (12)$$

$$T_{l+1}^a(n) = \frac{T_l(n)}{T_l(n) + 1} \quad (13)$$

$$G_T(n) = G(T_{l+1}^a) \quad (14)$$

where $G(\cdot)$ is gain control function as defined in [2]. The steps are followed by

$$A'''_l(n) = G_T(n) \times A_l(n) \quad (15)$$

$$A'_l(n) = |A'''_l(n)| \quad (16)$$

$$S'_l(n) = A'''_l(n) \times e^{i\Theta_l(n)} \quad (17)$$

$$y'_l = y'_{l-1} + \frac{re\left(\mathcal{F}^{-1}\left(S'_l(n)\right)\right)}{\frac{1}{\phi}} \quad (18)$$

Until iteration l reaches L. Here, y' is the output of TSNR. The final output of the TSNR is the real part of inverse of the Fourier transform $\mathcal{F}^{-1}(\cdot)$.

B. Experimental Setup

For experimental data, we used speech sample given by [5] which is a male voice with 8kHz sampling rated and 16b PCM. For noise model, we generate additive Gaussian noise with variance σ_N^2 obtained from SNR with signal power is the variance of original speech σ_S^2.

$$\sigma_N^2 = 10^{\log_{10}\sigma_S^2 - SNR} \quad (19)$$

We will add the noise to the sample speech to obtain noisy speech. We generate noise level for SNR 0 up to 0.5dB with increment 0.01dB. We choose low SNR to represent high additive noise. We apply Wiener Noise Reduction to the noisy speech to obtain filtered noisy speech. Finally, we will compare the Mean Square Error (MSE) of filtered speech and the original speech for every noise level. As comparison, we will compare the result of TSNR with Harmonic Regeneration Noise Reduction (HRNR).

III. Experimental Results and Discussion

As a result, we plot all the speech for noise level at SNR 0.5 dB. The following figure (Fig.1) is the result after normalized by its maximum value. The first-row speech is the original speech, we can see that there are non-speech parts that have low magnitude. The second-row wave is the random noise part. The wave contains no information, only random noise. The third-row wave is the noisy speech which is the original speech added by random noise. This wave is featured by non-speech parts where noise appears.

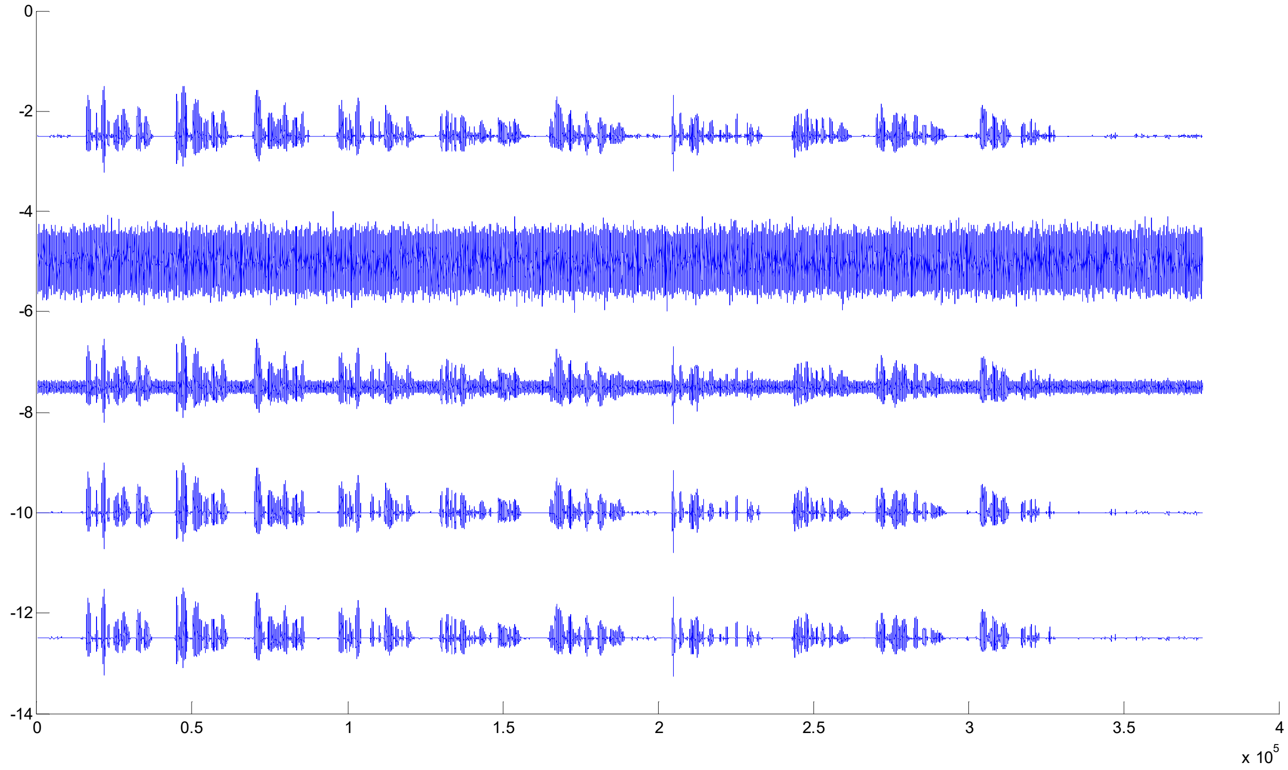

Fig. 1. Sound Wave.

The fourth and fifth waves are TSNR and HRNR wave filter result of the noisy speech. We can see that the noise has been decreased. The non-speech parts now appear better since the noisy part have been suppressed. However, the wave plots above do not quantitatively represent the improvement. The following figure shows the numerical performance of the filter by plotting the MSE for each low SNR as presented in previous section.

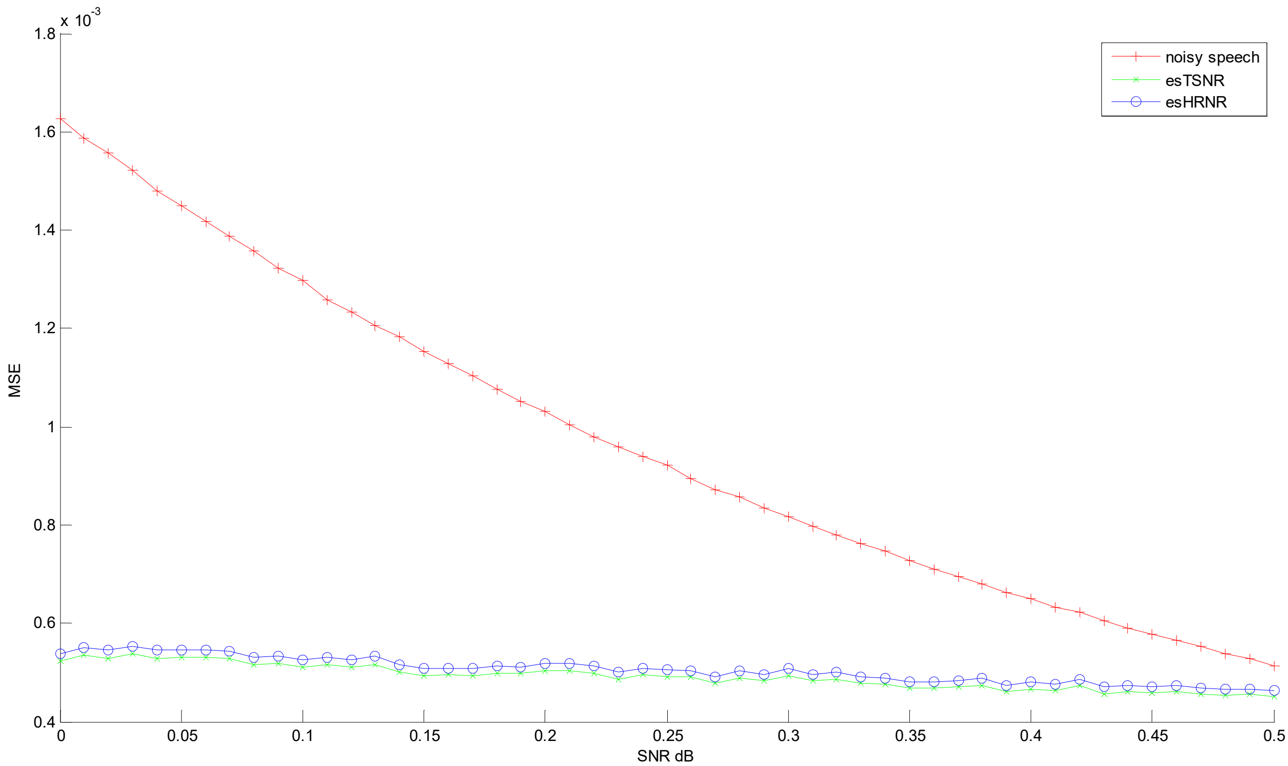

Fig. 2. MSE of the Result.

Figure 2 shows the error behavior as SNR increases. Both TSNR and HRNR show good result in noise reduction. Both techniques successfully reduce the noise when the noise level is very high. However, for all of our noise level, TSNR shows slightly better result in noise reduction than HRNR. Based on trend above, for low or very low noise level, TSNR and HRNR seem to be ineffective, since the noise reductions are not significant compared to noise level itself.

IV. Conclusion

COVID 19 has brought digital meeting and other multimedia communications becoming real. However, infrastructure for digital multimedia acquisition may not be available everywhere. Therefore, noise may distort the voice or other data. This paper evaluates the use of Wiener filter based denoising to enhance speech data quality. Enhancement of distorted speech by additive noise with only single observation has been done and still a challenging problem. TSNR successfully improve the speech in noisy environment up to order of 10^{-4} for low SNR <0.5 dB.

Acknowledgment

We would like to thanks to Prof. Azzedine Zerguine at KFUPM.

References

[1] A. N. Z. Rashed and M. S. F. Tabbour, "The Trade Off Between Different Modulation Schemes for Maximum Long Reach High Data Transmission Capacity Optical Orthogonal Frequency Division Multiplexing (OOFDM)," Wirel. Pers. Commun., 2018, doi: 10.1007/s11277-018-5690-9.

[2] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," IEEE Trans. Audio, Speech Lang. Process., 2006, doi: 10.1109/TASL.2006.872621.

[3] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," 1996, doi: 10.1109/icassp.1996.543199.

[4] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," IEEE Transactions on audio, speech, and language processing, vol. 14, pp. 1218-1234, 2016.

[5] B. R. Manju and M. R. Sneha, "ECG denoising using wiener filter and kalman filter," Procedia Computer Science, vol. 171, pp. 273-281, 2020.

[6] B. Somers, T. Francart, and A. Bertrand, "A generic EEG artifact removal algorithm based on the multi-channel Wiener filter," Journal of neural engineering, vol. 15, p. 036007, 2018.

[7] H. Salehi, J. Vahidi, T. Abdeljawad, A. Khan, and S. Y. B. Rad, "A SAR image despeckling method based on an extended adaptive wiener filter and extended guided filter," Remote Sensing, vol. 12, p. 2371, 2020.

[8] V. R. Balaji, J. S. Priya, J. R. D. Kumar, and S. P. Karthi, "Radial basis function neural networkbased speech enhancement system using SLANTLET transform through hybrid vector wiener filter," In Inventive Communication and Computational Technologies, pp. 711-723, Springer, Singapore, 2021.

[9] Baselice, F., Ferraioli, G., Ambrosanio, M., Pascazio, V., & Schirinzi, G. (2018). Enhanced Wiener filter for ultrasound image restoration. Computer methods and programs in biomedicine, 153, 71-81.

[10] A. Borowicz, "Using a multichannel Wiener filter to remove eye-blink artifacts from EEG data," Biomedical Signal Processing and Control, vol. 45, pp. 246-255, 2018.

[11] L. M. Dogariu, S. Ciochină, C. Paleologu, J. Benesty, and C. Oprea, "An iterative Wiener filter for the identification of multilinear forms," In 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), pp. 193-197, 2020.

[12] J. Vilà-Valls, D. Vivet, E. Chaumette, F. Vincent, and P. Closas, "Recursive linearly constrained Wiener filter for robust multi-channel signal processing," Signal Processing, vol. 167, p. 107291, 2020.

[13] A. Klos, M. S. Bos, R. Fernandes, and J. Bogusz, "Noise-dependent adaption of the Wiener filter for the GPS position time series," Mathematical Geosciences, vol. 51, pp. 53-73, 2019.

[14] H. H. Nuha and N. A. Suwastika. "Fractional fourier transform for decreasing seismic data lossy compression distortion." In 2015 3rd International Conference on Information and Communication Technology (ICoICT), pp. 590-593. IEEE, 2015.

[15] H. H. Nuha and M. Abido, "Firefly algorithm for log-likelihood optimization problem on speech recognition." In 2016 4th International Conference on Information and Communication Technology (ICoICT), pp. 1-6. IEEE, 2016.

[16] T. N. W., Siagian, H. H. Nuha, and R. Yasirandi, "Footstep Recognition Using Mel Frequency Cepstral Coefficients and Artificial Neural Network," Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), vol. 4, pp. 497-503, 2020.

[17] "American English. (n.d.)." http://www.voiptroubleshooter.com/open_speech/american.html (accessed May 09, 2015).