PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Wind noise removal from mixture with speech: Using Wiener filter and invariant frequency beamforming

Fan-Jie Kung[1]

[1]Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China,
sammykung2007@gmail.com

Abstract

In the paper, we attempt to design a system to help hearing-impaired hear clearly from a speaker in a noisy environment especially in high winds. First, we make an assumption that a listener always face directly to a speaker so that we can enhance the speech signal by beamforming technique. However, due to effects of high winds, we may also enhance the signal of the wind noise when the direction of the wind noise is the same as the speaker's. Therefore, we utilize frequency isolation, Wiener filter, in order to reduce wind noise of beamforming signals. For the beamforming technique, we utilize the delay and sum beamformer (DS BF). However, because the beam-width changes with frequency variance, some methods are used like minimum variance distortionless response (MVDR) and linearly minimum constrained variance (LMCV). Notwithstanding, these two methods are computation complexity. Hence, we use a multi-beamforming method to get the constant beam-width with low complexity. The data results show that this method of combination of multi-beamforming and revised Wiener filter improve the quality of sound of a speaker for the speed of wind at 4.0 m/s and it reduces the mean square error by approximately 77.639%, comparing with the distortion speech.

Keywords: Frequency isolation, Wiener filter, Beamforming

1 INTRODUCTION

Wind noise is a non-stationary signal and its mean and standard deviation change with time, unlike a white noise which belongs to a weak stationary signal and its mean and standard deviation are constant at all time points. Hence, in the paper, we analyse wind noise by short time Fourier transform (STFT) in a short period of time and approximately remove wind noise by frequency isolation concept, Wiener filter([1]-[3]).

Wiener filter has been used in image processing in order to restore the original image from distortion and this filter also is used in speech signal processing such as source separation. If we estimate the power spectra of wind noise and speech accurately, we can approximately separate the two sources. However, in real world, it is a little difficult to estimate the power spectra of wind noise and speech, especially when the two signals are mixed together. Therefore, in the paper, we measure the speed of wind by an anemometer to get a wind noise model and apply this model as noise to restore the clean speech from distortion by revised Wiener filter in section 2.3. However, this method places a limit of handling with high winds when the speed of wind is higher than 4.0 m/s. In order to deal with and improve this phenomenon, we apply a beamforming technique to get better performance.

Delay and sum beamforming has been used in microphone array system, telecommunication, radar and sonar because it brings effective results for reducing interference signal from a specific direction. Nevertheless, the beam-width of traditional delay and sum beamforming is narrower when the frequency of the incoming signal is higher. Hence, several invariant frequency beamforming methods have been proposed, such as minimum variance distortionless response (MVDR) and linearly minimum constrained variance (LMCV). However, the computation of these two methods are too complex to execute in real time when the number of sensors is high. In the paper, we apply multi-beamforming([4]-[7]) to reduce wind noise especially in high winds.

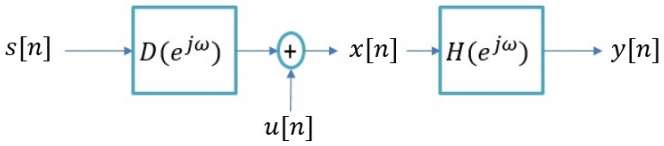

Figure 1. Illustration of Wiener filter and signal model

2 WIENER FILTER AND DATA COLLECTION

2.1 Wiener filter equation

Wiener filter can be derived from least mean square error estimation (LMSE) and its model is as shown in Fig. 1. In the paper, $s[n]$ represents a clean speech, $u[n]$ represents wind noise and $x[n]$ represents the mixed signal of a clean speech and wind noise and $y[n]$ is the results of Wiener filter. $D(e^{j\omega})$ represents the frequency response of distortion and $H(e^{j\omega})$ represents the frequency response of Wiener filter. In our experiment, we assume that the magnitude response of distortion system $|D(e^{j\omega})|$ is unity and phase response $\angle D(e^{j\omega})$ is zero so that a mixed signal of a clean speech and wind noise can be expressed as Eq. 1 and the relation of $y[n]$ and $x[n]$ is as Eq. 2.

$$x[n] = s[n] + u[n]. \tag{1}$$

$$y[n] = h[n] * x[n] = \sum_{k=0}^{p-1} h[k]x[n-k] = \mathbf{x}_n \mathbf{h}^T. \tag{2}$$

Where $h[n]$ represents the impulse response of Wiener filter, $\mathbf{x}_n = \{x[n], x[n-1], x[n-2]...x[n-p+1]\}$ is the sequence of the mixed signal of a clean speech and wind noise, where $n = 0, 1, 2...N-1$ and $\mathbf{h} = \{h[0], h[1], h[2]...h[p-1]\}$ represents the impulse response of Wiener filter so the error $e[n]$ can be estimated the vector as follows,

$$e[n] = s[n] - y[n] = s[n] - \mathbf{x}_n \mathbf{h}^T. \tag{3}$$

The average squared error function can be written as

$$E\{e[n]^2\} = E\{s[n] - \mathbf{x}_n \mathbf{h}^T\}^2 = E\{s[n]^2\} - 2\mathbf{h}^T E\{\mathbf{x}_n s[n]\} + \mathbf{h}^T E\{\mathbf{x}_n \mathbf{x}_n^T\}\mathbf{h} = \gamma_{ss}[0] - 2\mathbf{h}^T \gamma_{sx} + \mathbf{h}^T R_{\mathbf{xx}}\mathbf{h}, \tag{4}$$

where $R_{xx} = E\{\mathbf{x}_n \mathbf{x}_n^T\}$ is an auto-correlation matrix function of the mixed signal of speech and wind noise and $\gamma_{ss}[0]$ is an auto-correlation vector of speech signals and $\gamma_{sx} = E\{s[n]x[n]\}$ is a cross-correlation vector of speech signals and mixed signals of speech and wind noise. In order to achieve the minimum error, we take the partial derivation with respect to $\mathbf{h}$ in Eq. 4 and the following relation is derived,

$$\mathbf{h} = R_{xx}^{-1} \gamma_{sx}. \tag{5}$$

If we assume that wind noise is uncorrelated with speech signals, Eq. 5 can be rewritten as:

$$\mathbf{h} = (R_{ss} + R_{nn})^{-1} \gamma_{ss}, \tag{6}$$

and an equivalent frequency-domain representation of the Wiener filter can be expressed as follows,

$$H(e^{j\omega}) = \frac{|S(e^{j\omega})|^2}{|S(e^{j\omega})|^2 + |N(e^{j\omega})|^2}. \tag{7}$$

2.2 Wind noise analysis and preliminary trials

In the experiment, we recorded two groups of wind noise data, the first group of wind noise data was recorded in the schoolyard of National Tsing Hua University. The second group of wind noise data was recorded in the recording room by an electric fan and the recording device of these two groups is a built-in microphone of Sony cell phone. For the first group, we analyse the actual wind noise to attempt to know the properties of wind, for the second group, we attempt to record the similar wind noise with an approximately constant speed by an electric fan. At the beginning of the experiment, we recorded a clean speech and wind noise respectively and then mixed them together, the results are as shown in Fig. 2 to Fig. 4.

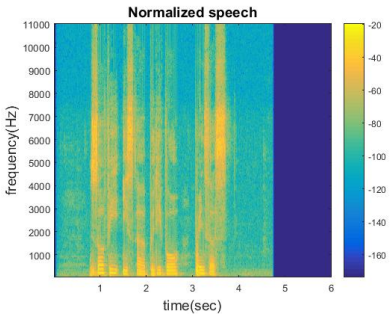

Figure 2. The spectrogram of the normalized speech

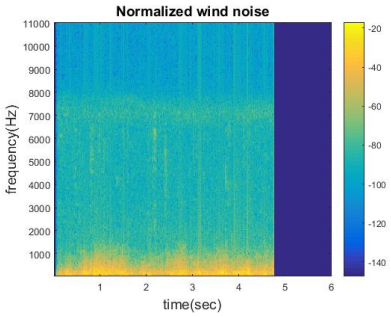

Figure 3. The spectrogram of the normalized wind noise

The context of the clean speech is "Sophie is a beautiful princess." and the Wiener filter result by Eq. 7 is as shown in Fig. 5.

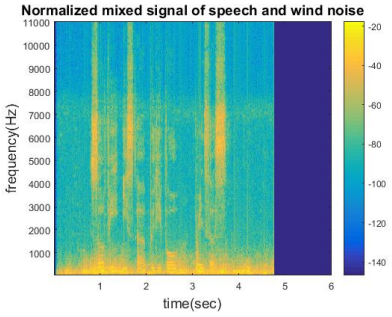

Figure 4. The spectrogram of the normalized mixed signal of speech and wind noise

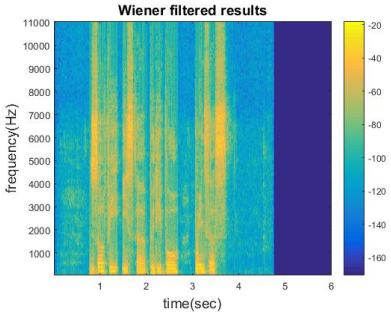

Figure 5. The spectrogram of Wiener filter result

From the above results, it is clear that we can approximately restore a clean speech from distortion by wind noise if we first know the spectra of a clean speech and wind noise. However, in real world, it is a little difficult to estimate the clean speech and wind noise when they are mixed together. In the next section, we introduce the revised Wiener filter to remove wind noise from a clean speech.

2.3 Revised Wiener filter

If we have the information of the spectrum of wind noise, we can approximately estimate the frequency response of a revised Wiener filter, as follows,

$$H_R(e^{j\omega}) = \frac{|X(e^{j\omega})|^2 - \alpha|N(e^{j\omega})|^2}{|X(e^{j\omega})|^2}. \tag{8}$$

Where $X(e^{j\omega})$ represents the spectrum of the mixed signal of a clean speech and wind noise and $N(e^{j\omega})$ represents the spectrum of wind noise. In this part, we recorded two kinds of speed of wind noise by an electric fan in a recording room, the first speed of wind by an electric fan is 2.4 m/s and the second speed of wind by a electric fan is 4.0 m/s. The wind speed corresponds to Beaufort scale 0 to 2. The Fig. 6 and Fig. 7 show the spectrogram of wind noise by an electric fan with different speeds: 2.4 m/s and 4.0 m/s respectively.

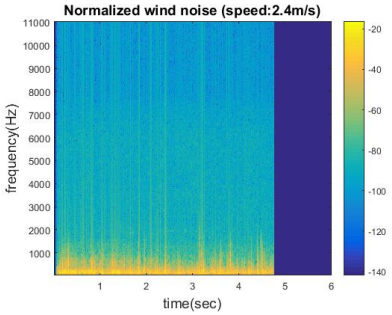

Figure 6. The spectrogram of the normalized wind noise (the speed of wind: 2.4 m/s)

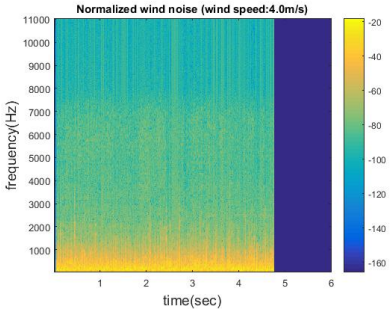

Figure 7. The spectrogram of the normalized wind noise (the speed of wind: 4.0 m/s)

It seems to be that when the speed of wind is higher, the average amplitude and frequency are higher, and the spectrogram of two mixed signals of a clean speech and wind noise are shown in Fig. 8 and Fig. 9.

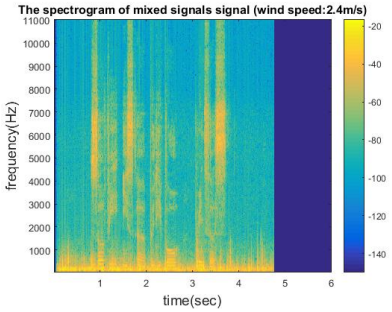

Figure 8. The spectrogram of mixed signals (speed of wind: 2.4 m/s)

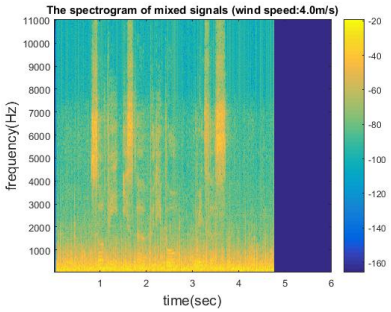

Figure 9. The spectrogram of mixed signals (speed of wind: 4.0 m/s)

In our experiment, we utilize revised Wiener filter to restore the clean speech from wind noise distortion and the value of α is 100.(due to the fact that we recently cannot predict the waveform of wind noise so that we approximately choose the high value of α) The spectrogram of two revised Wiener filter signals are shown in Fig. 10 and Fig. 11. From these results, in real world, it appears that it has a limitation of high winds source separation. Therefore, we combine the beamforming technique to solve this problem.

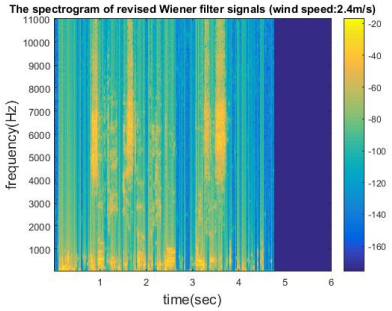

Figure 10. The spectrogram of revised Wiener filter signal (speed of wind:2.4m/s)

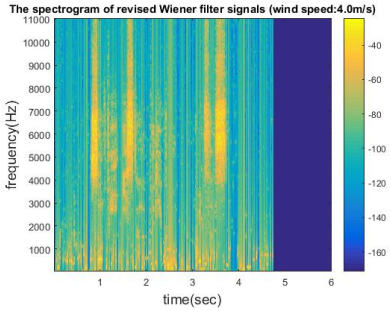

Figure 11. The spectrogram of revised Wiener filter signal (speed of wind:4.0m/s)

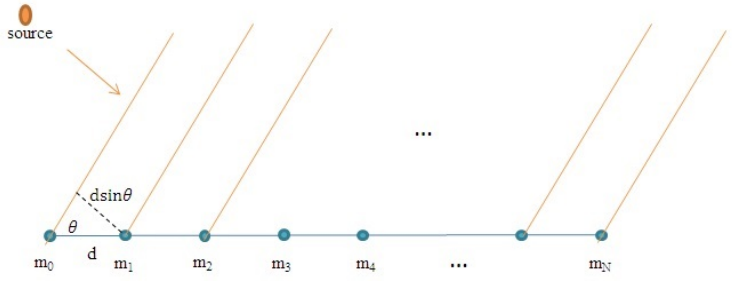

Figure 12. Illustration of uniform linear array

3 BEAMFORMING TECHNIQUES

3.1 Delay and sum beamforming with uniform linear array

The configuration of an uniform linear array (ULA) of microphones is shown in Fig. 12. If we have N microphones, the beampattern can be expressed as follows,

$$A_n(\theta) = \sum_{n=0}^{N-1} e^{-j2\pi f \tau_n}. \tag{9}$$

Where, $\tau_n = \frac{nd\sin\theta}{c}$ is the time delay for each microphone (so the index delay is $\tau_n \times f_s$, f_s represents the sampling frequency), d is the distance between adjacent microphones. c is the speed of sound. By the beam-width definition, the beam-width can be expressed as $\theta_{BW} = 2\arcsin\frac{c}{Ndf}$. Clearly, the beam-width is dependent on the number of microphones, the distance of microphone array and the frequency of the source. If we fix the number of microphones and the distance d, then θ_{BW} decreases when f increases. The frequency of the source is higher, the beam-width is narrower. In contrast, the source of the frequency is lower, the beam-width is wider. The microphone array beam-width is as shown in Fig. 13 for the number of microphone array is 8 and the distance between each microphone is 0.2m.

3.2 Multi-beamforming

The multi-beamforming method makes the beam-width constant over a wide frequency range. That is, $\theta_{BW} = 2\arcsin\left(\frac{c}{Ndf} + \frac{\tau(\omega)}{\tau_0}\right)$ is constant from the specific frequency range. From the above equation, the phase delay

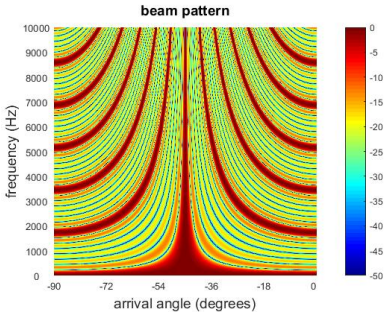

Figure 13. The beam-pattern of traditional ULA system

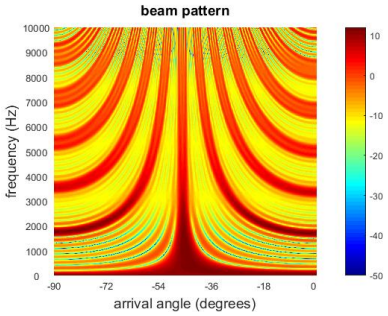

Figure 14. The beam-pattern of multi-beamforming system

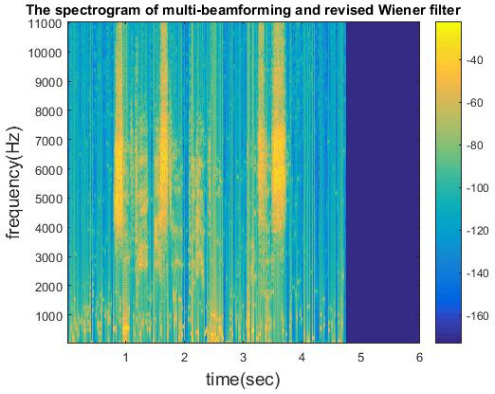

Figure 15. The spectrogram of multi-beamforming and revised Wiener filter

$\tau(\omega)$ is $\frac{1}{Nf}(\frac{\omega}{\omega_0}-1)$, so that the phase function $f(\omega)$ is $\frac{2\pi}{N}(1-\frac{\omega}{\omega_0})$, and the results is shown as Fig14, where $\omega_0 = 2\pi(1000)$ and in this multi-beamforming method experiment, the formula is as shown in Eq. 10.

$$A_n(\omega,\theta) = \sum_{i=0}^{r-1}\sum_{n=0}^{N-1} a_n e^{-jn\omega\tau_0\sin\theta} e^{jnf(\omega)\frac{i}{r}}. \tag{10}$$

Where r is $N/2$, $\tau_0 = \frac{d}{c}$ and a_n represents the frequency response of source by n^{th} channel and N is the number of microphone in an array system. In our experiment, we make an assumption that speakers always talk with each other face to face so that the value of θ is zero and Eq. 10 can be expressed as follows,

$$A_n(\omega,\theta) = \sum_{i=0}^{r-1}\sum_{n=0}^{N-1} a_n(1) e^{jnf(\omega)\frac{i}{r}}. \tag{11}$$

4 RESULTS AND DISCUSSION

4.1 Results

The result of multi-beamforming and revised Wiener filter by Eq. 11 is as shown in Fig. 15, under the condition of the speed of wind is 4.0 m/s, comparing with the result of revised Wiener filter (Fig. 11) and the distorted speech (Fig. 9), the results of mean square error (MSE) are shown in Table. 1.

The results of a mean square error table is as follows.

Table 1. The results of mean square error (wind speed: 4.0 m/s)

The type of audio source	distorted speech	revised Wiener filter	multi-beamforming and revised Wiener filter
MSE	0.0322	0.0077	0.0072

4.2 Discussion

We initially assumed that speakers always talk with each other face to face, applying the combination of multi-beamforming and revised Wiener filter to remove wind noise from speech. This study uses data from the recording room by an electric fan, to get approximately a constant wind speed, particularly for 4.0 m/s, measured by an anemometer and a recording device, a built-in microphone of a Sony cell phone. The results show that the performance of the combination of multi-beamforming and revised Wiener filter is better than that of revised Wiener filter only, reducing the mean square error by approximately 6.49%. In comparison with the distorted speech, the multi-beamforming and revised Wiener filter can reduce the mean square error by approximately 77.639%. Reliance on these measures should be tempered, for the presumed notion that speakers always talk with each other face to face and the man-made wind noise. Overall, it appears that under these conditions, we can improve the performance of distorted speech, suggesting that we can produce, in the future, the waveform of wind noise according to their speed, in order to design a better framework, handling with wind noise separation from a clean speech.

5 CONCLUSIONS

Our findings show that we can utilize multi-beamforming and revised Wiener filter to improve the performance that compares with the distorted speech, especially in high winds of approximately at 4.0 m/s.

REFERENCES

[1] Saeed V. Vaseghi. "Advanced Digital Signal Processing and Noise Reduction," John Wiley and Sons Ltd, second edition, 2000.

[2] Sascha. Spors. "Digital Signal Processing," Universität Rostock (Germany), 2015.

[3] C.M. Nelke. "Wind noise reduction-signal processing concepts," RWTH Aachen University (Germany), 2016.

[4] M. Goodwin., G. W. Elko. "CONSTANT BEAMWIDTH BEAMFORMING," ICASSP, Aug 1993.

[5] R. Smith. "Constant beamwidth receiving arrays for broad band sonar systems," Acoustic, vol.23, 1970, pp21-26.

[6] E. Hixson., k. Au. "Broadband constant beamwidth acoustical arrays," Technical Report 19, Acoustics Research Lab, U.T.Austin, 1970.

[7] D. Tucker. "Arrays with constant beam-width over a wide frequency range," Nature, vol.180, Sept 1957, pp496-497.