Research Journal of Applied Sciences, Engineering and Technology 6(5): 919-923, 2013
DOI:10.19026/rjaset.6.4142
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 30, 2012 Accepted: January 03, 2013 Published: June 25, 2013

Research Article

Improvement for Speech Signal based on Post Wiener Filter and Adjustable Beam-Former

[1, 2]Xiaorong Tong and [1]Xiangfeng Meng
[1]College of Mathematics and Information Science, Weinan Normal University, Weinan, Shaanxi, China
[2]Department of Electronics Engineering, Northwestern Polyteehnieal University, Xi'an, China

Abstract: In this study, a two-stage filter structure is introduced for speech enhancement. The first stage is an adjustable filter and sum beam-former with four-microphone array. The control of beam-forming filter is realized by adjusting only a single control variable. Different from the adaptive beam-forming filter, the proposed filter structure does not bring to any adaptive error noise, thus, it also does not bring the trouble to the second stage of the speech signal processing. The second stage of the proposed filter is a Wiener filter. The estimation of signal's power spectrum for Wiener filter is realized by cross-correlation between primary outputs of two adjacent directional beams. This estimation is based on the assumption that the noise outputs of the two adjacent directional beams come from two independent noise source but the speech outputs come from the same speech source. The simulation results shown that the proposed algorithm can improve the Signal-Noise-Ratio (SNR) about 6 dB.

Keywords: Adjustable filter, beam forming, post wiener filter, speech enhancement

INTRODUCTION

Mobile carriers are intimately aware of the role that voice quality plays in customer retention. One of the primary factors affecting voice quality is environmental noise and so any mean of suppressing noise provides a potential differentiator for handset manufacturers. Until recently, noise suppression technology focused on reducing slow-changing stationary noise sources. Such as a loud fan in the background-stationary noise can be recognized and effectively subtracted through conventional techniques. However, many noise sources, known as non-stationary noise, are fast-changing and not being suppressed. This non-stationary noise is normal in life, such as a person talking, background music, or keyboard typing. As a result, subscribers are unable to reliably use their handsets on busy streets, in crowded restaurants, or even at home. Therefore, next generation noise suppression technology use multiple microphones to more accurately identify, locate and group noise and speech sources which is little possible with a single microphone.

The beam-forming systems applied to microphone array application are used commonly for improving the quality of a received signal. Typical application can be found not only in the handset system mentioned above, but also the hands-free situation when talking to phone while driving cars. The most frequently studied beam-forming methods are focus on adaptive beam-former like the Frost's linearly constrained adaptive beam-former (Frost, 1972). The problem of these adaptive beam-formers is that they may cause unpredictable distortion of desired signal (Grenier, 1993). This is not easy to minimize by some adaptive interference cancelers. However, in Goulding and Bird (1990), Affes and Grenier (1994) and Bitzer et *al*. (1999), it is proved that the adaptive interference canceller cannot perform better than an optimized constant beam-former for a given application, such as a car cabin environment. Another advantage of constant beam-forming is that the filtering performance is deterministic. It is favorable for the following process, for example, the estimation of power spectrum. Due to the two reasons mentioned above, adjustable filter and sum beam-former (Kajala and Hamalainen, 2001) is used as the first stage of filter. The optimized filters for different view directions were computed in advance. In order to support the view steering directions, the collection of optimal filter coefficients would have to be stored in a memory and continuously updated by retrieving filter coefficients from the memory.

The second stage of filter is Wiener filter. The proposed algorithm for the estimation of the ideal signal's power system is computed by short-term measurement of the autocorrelation and cross-correlation function of the microphones (Zelinski, 1988).

METHODOLOGY

Filter and sum beam-former:

Input signal: The microphone array is composed by four omnidirectional microphones. The input signal

Corresponding Author: Xiaorong Tong, College of Mathematics and Information Science, Weinan Normal University, Weinan, Shaanxi, China

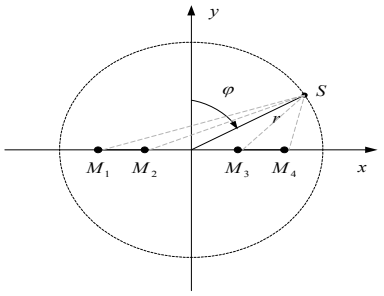

Fig. 1: Linear array of four microphones Mi and one signal point source S

sources are all point sources emitting spherical waves. We can now write the output of an ideal omnidirectional microphone at location M as:

$$x_i(t) = \frac{s(t - r_i / c)}{r_i} \quad (1)$$

where,

c = The speed of sound
$s(t)$ = The wave equation of point source
r_i = The distance from the source location S to the microphone position M

In this study, we use 1-D signal as an example. Thus, a source location S can be defined by only using direction, when r is already known. The array is illustrated in Fig. 1.

We assume *S* the source equation is a normalized cosine wave with frequency ω_0, the location of microphones is m_i, $1 \le i \le N$. According to Eq. (1), we get:

$$x_i(t) = \frac{\cos(w_0(t - r_i / c))}{r_i}$$
$$r_i = \sqrt{(r \sin\varphi - m_i)^2 + (r \cos\varphi)^2} \quad (2)$$

Filter and sum beam-forming filter structure: The filter and sum beam-forming filter consist of M FIR filters with the Length L, as shown in Fig. 2.

where, h_{ik} is the filter coefficient, according to Eq. (1) and some transfers, we can obtain:

$$y(n) = f_{ML}(n) \cos w_0 n \quad (3)$$

where, $f_{ML}(n)$ is the impulse response of the filter-and-sum beam-forming filter:

$$f_m(n) = \sum_{i=1}^{M} \sum_{k=0}^{L-1} \frac{h_{ik}}{r_i} \delta(n - r_i / c - k) \quad (4)$$

After implementing Fourier transform, it is desirable:

$$Y(w) = F_{ML}(w) \exp(j w_0) \quad (5)$$

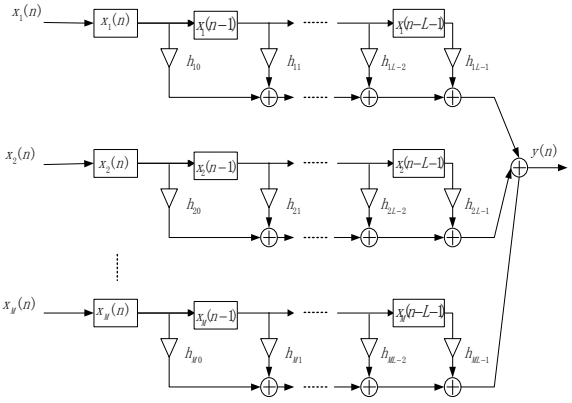

Fig. 2: Filter and sum beam-forming filter structure

where,

$$F_{ML}(w) \sum_{i=1}^{M} \sum_{k=0}^{L-1} \frac{h_{ik}}{\sqrt{(r \sin\varphi - m_i)^2 + (r \cos\varphi)^2}}$$

$$\exp(-jw\,(k + \sqrt{(r \sin\varphi - m_i)^2 + (r \cos\varphi)^2}\,/c)) \quad (6)$$

Then, we can expand to N input signals with the location i, which has Fourier transform $S_i(w)$, $1 \le i \le N$.

$$Y(w) = \sum_{i=1}^{N} F_{ML}(w, \varphi_i) S_i(w) \quad (7)$$

One of the design targets is to optimize the filter response $F_{ML}(w, \varphi)$ to provide flat magnitude response in a desired view direction φ_{des}. Another criterion is to have a flat spectrum in any other direction as well. The third aim to enable steering of the desired view direction φ_{des} in any direction $0^0 \le \varphi_{des} \le 180°$. Figure 3 illustrate the target response $|F_{ML\varphi des}(w, \varphi)|$ used for filter optimization and the corresponding direction of φ_{des}.

If we want to get a resolution with 10°, that is 18 sets of $h_{ik}(\varphi_{des})$. Due to the relation between π-φ_{des} and φ_{des}, by switching the order of microphones coefficient $h_{ik}(\varphi_{des})$, we can easily get $h_{ik}(\pi$-$\varphi_{des})$. Therefore the memory for $h_{ik}(\varphi_{des})$ would be reduced by half numbers.

Post wiener filter: After the beam-former with the ideal magnitude response $|F_{ML\varphi des}(w, \varphi)|$ and zero phase delay, we can get desired point source signal function (located in φ_{des}):

$$Y(w) = \sum_{n=1}^{N} F_{ML\varphi des}(w, \varphi_i) X_i(w) \approx F_{ML\varphi des}(w, \varphi_{des}) X_{des}(w) = c X_{des}(w) \quad (8)$$

In a favorable situation for Wiener filter, the aperture of microphone array is big enough that speech source is not a point source. It covers numbers of points. The adjacent outputs X_i and X_j, with viewing locations of φ_i and φ_j, all in the area of speech

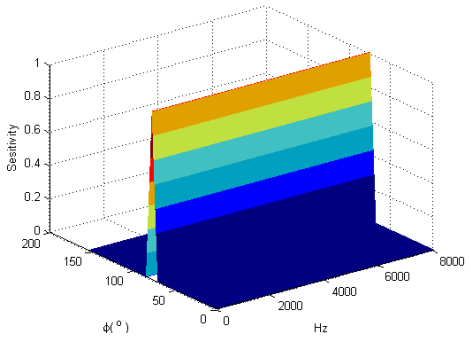

Fig. 3: Design target response $|F_{ML\varphi des}(w, \varphi)|$ of $\varphi_{des} = 60°$

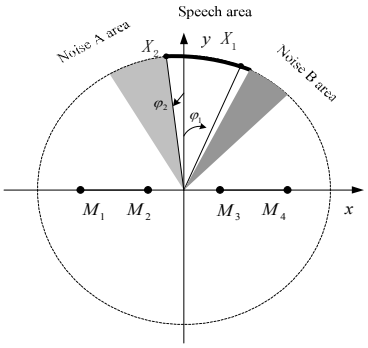

Fig. 4: Simulation environment

source, both have the speech signal spectrum information and noise:

$$X_i(w) = cS(w) + N_i(w)$$
$$X_j(w) = cS(w) + N_j(w) \quad (9)$$

In the time domain:

$$x_i(n) = cs(n) + n_i(n)$$
$$x_j(n) = cs(n) + n_j(n) \quad (10)$$

Let's discuss the simplest situation that $n_i(n)$ and $n_j(n)$ come from two independent noise sources. The estimation of signal power spectrum will be as follows:

$$F_{ss}(w) = F_{ij}(w) \quad (11)$$

where, $\Phi_{ij}(\omega)$ is the cross power spectrum of output $x_i(n)$ and $x_j(n)$
Then, the Wiener filter:

$$W(w) = \frac{F_{ss}(w)}{F_{ss}(w) + F_{nn}(w)} \quad (12)$$

However, this situation does fit the reality very well. Problem occurs when noise outputs from the two adjacent locations relate with each other, especially in the low frequency. Then the estimation of speech power spectrum is not going to be easy. There are some solutions like the combination of autocorrelation and cross-correlation for the estimation of power spectrum (Meyer and Simmer, 1997; McCowan and Bourland, 2002).

Simulation: We run a 4-element linear microphone array of omnidirectional transducers for beam-forming filter structure shown in Fig. 2. Select 5 cm sensor spacing to avoid spatial aliasing on the highest frequency 3.4 kHz for narrowband speech signal.

Simulation environment: We create a simulation environment describe in the Fig. 4.

Where $\varphi_1 = 20°$, $\varphi_2 = -10°$ are two viewing directions of beam-former, which is already determined. Also r = 0.4 m is one of variables determined early. ∠Speecharea = 40°, ∠NoiseAarea = 40° and ∠NoiseBarea = 40°, the three variables define covering areas of speech, noise A and noise B. Noise A is babble noise. The source of this babble is 100 people speaking in a canteen. The room radius is over two meters; therefore, individual voices are slightly audible. Noise B is vehicle interior noise. They construct an ideal simulation of real environment.

Design criteria: We develop a method to optimize the direction sensitivity of a filter-and sum beam-former. The directivity of microphone array is optimized by adjusting the filter coefficients to minimize the mean square error between the desired and actual response of the beam-former (Kajala and Hamalainen, 2001).
Define the MSE as:

$$MSE = \sum_{w,\varphi \in \Omega_{w,f}} \frac{\left(|F_{ML}(w, \varphi)| - |F_{MLdes}(w, \varphi)|\right)^2}{|\Omega_{w,\varphi}|} \quad (13)$$

where,
$|F_{ML}(w, \varphi)|$: The filter response in the frequency domain
$\Omega_{\omega,\varphi}$: The sets of ω, φ, we need to compute $F_{ML}(w, \varphi)$ and $F_{MLdes}(w, \varphi)$

In order to meet the design criteria, we optimize filter parameters h_{ik} for M = 4 and L = 20.

SIMULATION RESULTS

The following figure illustrates the performance for filter and sum beam-former response, the speech as' this is the VOA special English' and 'independent' respectively (Fig. 5).

The performance of outputs is measured by ALSD (average log-spectrum Distance), defined as:

$$D(w) = \frac{1}{P}\sum_{i=0}^{P-1} \left| ln(X_i(w)) - ln(\hat{X}_i(w)) \right| \quad (14)$$

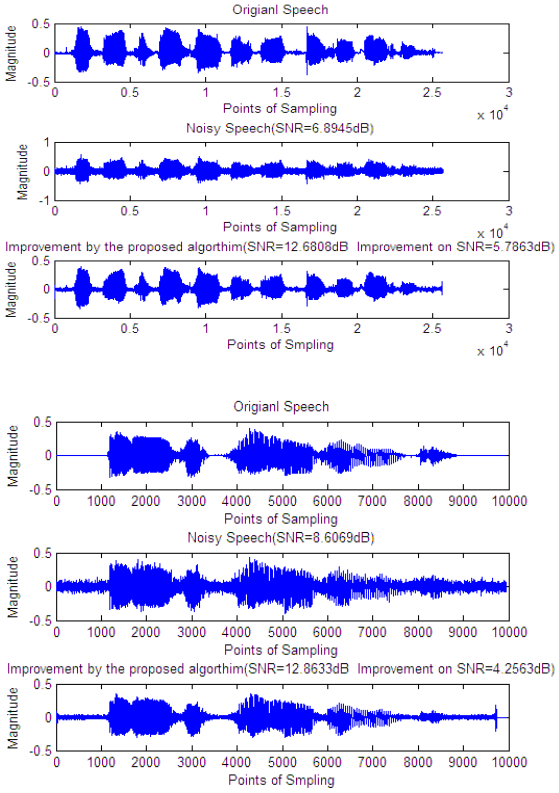

Fig. 5: The speech of 'this is the VOA special English' (top) and 'independent' (bottom)

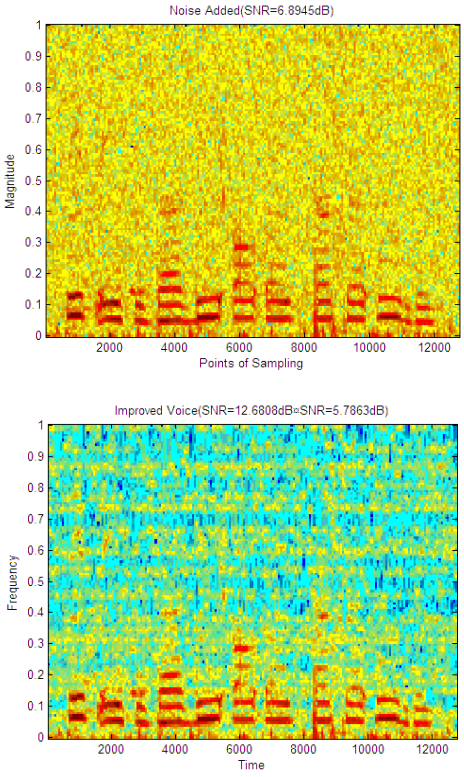

Fig. 6: Spectrograms of the noisy speech signal (top) and the de-noising speech signal (bottom)

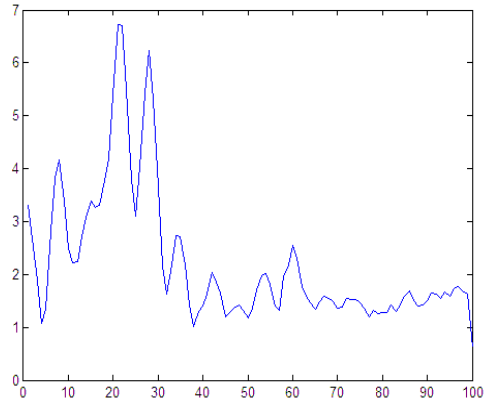

Fig. 7: ALDS measure of the distortion

The spectrograms of the noisy speech signal and the de-noising speech signal by the proposed algorithm are also shown in the Fig. 6.

In Fig. 7, it has shown the performance of signal outputs which was measured by Average Log-Spectrum Distance (ALSD). Combined Fig. 6 we can see that the output of speech signal has a low ALDS and an excellent performance and the SNR increased about 6 dB by the proposed algorithm.

CONCLUSION

In this study, we introduced a newly two-stage filter by the microphone array processing, which allows suppressing the unpleasant noise around the speaker effectively and providing a relatively clear environment. It used both the spatial and time processing methods to reach this goal. The simulation results have shown that the proposed method has an excellent ability to suppress the noise in noisy signal and the speech has a considerable improvement.

ACKNOWLEDGMENT

The authors are very grateful to the anonymous referees for their careful reading, helpful comments. This study was supported by Weinan Science and Technology Initiatives Fund program (2012KYJ-8), and Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 12JK0745).

REFERENCES

Affes, S. and Y. Grenier, 1994. The adaptive beam-formers for speech acquisition in cars. Proceeding of the International Conference on Signal Processing Applications and Technology, pp: 154-159.

Bitzer, J., K.U. Simmer and K.D. Kammeyer, 1999. Theoretical noise reduction limits of the generalized side lobe canceller (GSC) for speech enhancement. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 5.

Frost, O.L., 1972. An algorithm for linearly constrained adaptive array processing. Proc. IEEE, 60(8): 926-935.

Goulding, M. and J.S. Bird, 1990. Speech Enhancement for mobile telephony. IEEE T. Veh. Technol., 39(4): 316-326.

Grenier, Y., 1993. A microphone array for car environments. Speech Commun., 12: 25-39.

Kajala, M. and M. Hamalainen, 2001. Filter-and-sum beam-former with adjustable filter characteristics. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 5.

McCowan, I.A. and H. Bourland, 2002. Microphone array post-filter for diffuse noise field. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 1.

Meyer, J. and K.U. Simmer, 1997. Multichannel speech enhancement in car environment using Wiener filtering and spectral subtraction. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 2.

Zelinski, R., 1988. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 5.