A MULTIFRAME PARAMETRIC WIENER FILTER FOR ACOUSTIC ECHO SUPPRESSION

Hai Huang[1,2], Christian Hofmann[2], Walter Kellermann[2], Jingdong Chen[1], and Jacob Benesty[3]

[1] IAIC Research Center
Northwestern Polytechnical University
127 Youyi West Road
Xi'an, Shaanxi 710072, China

[2] Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstraße 7
91058 Erlangen, Germany

[3] INRS-EMT
University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

ABSTRACT

Acoustic echo arises due to the acoustic coupling between the loudspeaker and the microphone in a full-duplex voice communication device. How to reduce or eliminate echo has been an important problem in voice communications. This paper deals with this problem in the short-time Fourier transform (STFT) domain. An approach to acoustic echo suppression (AES) is developed, which uses a linear filter in every STFT subband. A non-parametric variable step-size (NPVSS) normalized-least-mean-square (NLMS) algorithm is used to estimate the echo signal in every STFT bin and time frame. We then define the desired output signal as the sum of the desired near-end speech and a residual echo component. A parametric Wiener filter (PWF) is subsequently derived by minimizing the near-end speech distortion with a constraint on the residual echo. There are two parameters in this PWF, which can be adjusted to control the compromise between the amount of echo attenuation and the degree of near-end speech distortion, and the level of the residual echo remaining at the output. These two parameters can be tuned to optimize the perceptual quality of the echo-reduced signal. Experimental results demonstrate that the developed PWF can effectively attenuate the undesired echo and estimate the near-end signal with little distortion.

Index Terms— Acoustic echo suppression (AES), interframe correlation, parametric Wiener filter (PWF), tradeoff filter, residual echo control.

1. INTRODUCTION

Controlling the adverse effect of acoustic echo, which results from the acoustic coupling between a loudspeaker and a microphone, is a well-known problem in acoustic signal processing for a full-duplex hands-free telecommunication systems. In the literature, two different techniques, namely acoustic echo cancellation (AEC) and acoustic echo suppression (AES), have been employed to eliminate or reduce the undesired echoes.

AEC is a linear filtering approach, which assumes that the echo path of the loudspeaker-enclosure-microphone (LEM) system is linear and can be identified with an adaptive filter. The adaptive filter yields an estimate of the echo signal, which is then subtracted from the microphone signal [1]–[3]. Ideally, AEC can eliminate echoes without introducing any distortion to the desired near-end signal [4]. However, time-varying acoustic paths [5], undermodeling [6], and nonlinearities in low-cost or miniaturized audio hardware [7] may hamper good identification of the echo path, which limits the performance of AEC. Furthermore, double-talk [2], [3] (simultaneously active far-end and near-end speakers) is another factor that affects the AEC performance. In many applications, the amount of echo attenuation with AEC may not be sufficient. In this case, a postfilter is generally used in conjunction with AEC to further remove the residual echo [2], [3], [8].

Alternatively, AES can be used to attenuate echo [9]–[17], which acts similarly to the above-mentioned postfilter, but without the AEC part. Generally, AES achieves echo attenuation using a parametric spectral modification algorithm to the microphone signal in the frequency domain, which is a technique originally developed for noise reduction [18]. An early study on AES can be found in [9]. Then, a perceptually motivated approach termed perceptual acoustic echo suppressor (PAES) was proposed in [11]. This approach estimates the spectral envelope of the echo signal while taking into account the frequency-selective properties of the human auditory system. Other methods model the acoustic echo path by means of subband filter gains, which lead to an overall delay and coloration of the short-time spectrum at each frequency of the short-time spectra [12]–[15]. A soft decision method was developed in [16] to compute the AES filter gain.

The major advantage of AES over AEC is its robustness to nonlinearity and double-talk [10]. The computational complexity of AES is generally also lower in comparison of that of AEC. However, AES introduces near-end speech distortion. How to control the amount of speech distortion is an open problem in AES and motivates this work.

In this paper, we develop an AES approach by considering the interframe correlation in the short-time Fourier transform (STFT) domain. The overall scheme is illustrated in Fig. 1. First, both the loudspeaker and the microphone signals are partitioned into small, overlapping frames with a typical frame length between 10 and 40 ms. Later on, each windowed frame is transformed into the STFT domain. Next, the echo component is estimated by modeling the echo path in each frequency bin and time frame. An FIR filter is designed in each subband based on the estimated echo and microphone signals. This FIR filter is then applied to the microphone signal such that the echo components are suppressed while maintaining the desired near-end signal. Finally, the echo suppressed signal is reconstructed by transforming the echo-reduced STFT coefficients back to the time domain with the inverse STFT.

2. SIGNAL MODEL AND PROBLEM FORMULATION

Let us consider the conventional signal model shown in Fig. 1, where acoustic echoes are generated from the linear coupling between a loudspeaker and a microphone [6]. Therein, the discrete-time microphone signal at time index n can be written as

$$d(n) = u(n) + g(n) * x(n) = u(n) + y(n), \tag{1}$$

where $u(n)$ is the near-end signal, $x(n)$ is the loudspeaker (or far-end) signal, $g(n)$ is the impulse response from the loudspeaker to the microphone, and $y(n)$ is the echo signal. All signals in (1) are considered to be real-valued, zero mean, and broadband, and we assume

This work is supported in part by the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005. The work of H. Huang is supported in part by the China Scholarship Council.

that $u(n)$ and $y(n)$ are uncorrelated.

Using the STFT, the signal model given in (1) can be expressed in the time-frequency domain as

$$D(k,m) = U(k,m) + Y(k,m), \tag{2}$$

where $D(k,m)$, $U(k,m)$, and $Y(k,m)$ are the STFTs of $d(n)$, $u(n)$, and $y(n)$, respectively, at the frequency bin $k \in \{0, 1, \dots, K-1\}$ and the time frame m.

For the purpose of considering interframe signal correlation later on, we concatenate L consecutive frames of the microphone signal at the frequency bin k to a vector:

$$\begin{aligned}\mathbf{d}(k,m) &= [D(k,m)\ D(k,m-1)\ \cdots\ D(k,m-L+1)]^T \\ &= \mathbf{u}(k,m) + \mathbf{y}(k,m),\end{aligned} \tag{3}$$

where the superscript ^{T} denotes transposition, and $\mathbf{u}(k,m)$ and $\mathbf{y}(k,m)$ are defined in an analogous way to $\mathbf{d}(k,m)$. Since $\mathbf{u}(k,m)$ and $\mathbf{y}(k,m)$ are uncorrelated by assumption, the correlation matrix (of size $L \times L$) of $\mathbf{d}(k,m)$ is

$$\begin{aligned}\mathbf{\Phi_d}(k,m) &\triangleq \mathcal{E}\left\{\mathbf{d}(k,m)\mathbf{d}^H(k,m)\right\} \\ &= \mathbf{\Phi_u}(k,m) + \mathbf{\Phi_y}(k,m),\end{aligned} \tag{4}$$

where $\mathcal{E}\{\cdot\}$ denotes mathematical expectation, the superscript ^{H} is the transpose-conjugation operator, and $\mathbf{\Phi_u}(k,m)$ and $\mathbf{\Phi_y}(k,m)$ are the correlation matrices of $\mathbf{u}(k,m)$ and $\mathbf{y}(k,m)$, respectively.

Generally in echo reduction, the objective is to extract the desired near-end speech component $U(k,m)$ and to attenuate the echo component $Y(k,m)$. As mentioned in the introduction, traditional AES may introduce musical noise, which is unpleasant to listen. One way to limit the effect of musical noise is to control the amount of the residual echo, which can mask the musical noise. To this end, we define a target signal as the sum of the desired near-end speech and a target residual echo component as

$$T(k,m) \triangleq \mathbf{i}_1^H \mathbf{u}(k,m) + \alpha \mathbf{i}_1^H \mathbf{y}(k,m), \tag{5}$$

where the parameter $0 \le \alpha \le 1$ controls how much residual echo remains in the target signal and $\mathbf{i}_1 = [1\ 0\ \cdots\ 0]^T$ selects only signal components from frame m.

We aim at obtaining an estimate $\widehat{U}(k,m)$ of the defined target signal in (5) by applying an FIR filter $\mathbf{h}(k,m)$ to the microphone signal in every subband as illustrated in Fig. 1 [17], [19], i.e.,

$$\begin{aligned}\widehat{U}(k,m) &= \mathbf{h}^H(k,m)\mathbf{d}(k,m) \\ &= U_{\mathrm{f}}(k,m) + Y_{\mathrm{re}}(k,m),\end{aligned} \tag{6}$$

where $\mathbf{h}(k,m)$ is a complex-valued filter of length L,

$$U_{\mathrm{f}}(k,m) = \mathbf{h}^H(k,m)\mathbf{u}(k,m) \tag{7}$$

is a filtered version of the desired near-end subband signal, and

$$Y_{\mathrm{re}}(k,m) = \mathbf{h}^H(k,m)\mathbf{y}(k,m) \tag{8}$$

is the residual echo, which is uncorrelated with $U_{\mathrm{f}}(k,m)$.

3. THE PWF FOR ECHO SUPPRESSION

3.1. Derivation of the Parametric Wiener Filter

We consider using the parametric Wiener filter in our problem with the newly defined target signal in (5). The error signal between the estimated and target signals at the frequency bin k and time frame m is

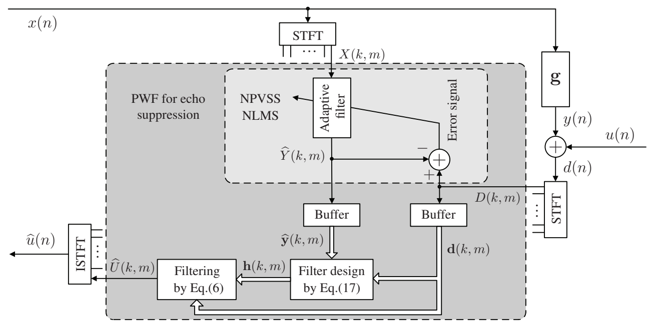

Fig. 1. Block diagram of echo suppression by applying an FIR filter in the STFT domain.

$$E(k,m) \triangleq \widehat{U}(k,m) - T(k,m). \tag{9}$$

Substituting (5) and (6) into (9), we can rewrite the error signal as

$$E(k,m) = E_{\mathrm{sd}}(k,m) + E_{\mathrm{re}}(k,m), \tag{10}$$

where

$$\begin{aligned}E_{\mathrm{sd}}(k,m) &\triangleq \mathbf{h}^H(k,m)\mathbf{u}(k,m) - \mathbf{i}_1^H \mathbf{u}(k,m) \\ &= [\mathbf{h}(k,m) - \mathbf{i}_1]^H \mathbf{u}(k,m)\end{aligned} \tag{11}$$

is the speech distortion due to filtering and

$$\begin{aligned}E_{\mathrm{re}}(k,m) &\triangleq \mathbf{h}^H(k,m)\mathbf{y}(k,m) - \alpha\mathbf{i}_1^H \mathbf{y}(k,m) \\ &= [\mathbf{h}(k,m) - \alpha\mathbf{i}_1]^H \mathbf{y}(k,m)\end{aligned} \tag{12}$$

represents the error between the filtered echo and the target residual echo.

Having defined the error signal, we can now write the subband mean-squared error (MSE) criterion:

$$\begin{aligned}J\left[\mathbf{h}(k,m)\right] &\triangleq \mathcal{E}\left\{|E(k,m)|^2\right\} \\ &= J_{\mathrm{sd}}\left[\mathbf{h}(k,m)\right] + J_{\mathrm{re}}\left[\mathbf{h}(k,m)\right]\end{aligned} \tag{13}$$

where

$$\begin{aligned}J_{\mathrm{sd}}\left[\mathbf{h}(k,m)\right] &\triangleq \mathcal{E}\left\{|E_{\mathrm{sd}}(k,m)|^2\right\} \\ &= \mathcal{E}\left\{\left|[\mathbf{h}(k,m) - \mathbf{i}_1]^H \mathbf{u}(k,m)\right|^2\right\},\end{aligned} \tag{14}$$

$$\begin{aligned}J_{\mathrm{re}}\left[\mathbf{h}(k,m)\right] &\triangleq \mathcal{E}\left\{|E_{\mathrm{re}}(k,m)|^2\right\} \\ &= \mathcal{E}\left\{\left|[\mathbf{h}(k,m) - \alpha\mathbf{i}_1]^H \mathbf{y}(k,m)\right|^2\right\}.\end{aligned} \tag{15}$$

Given the above MSEs, we can now derive optimal filters by minimizing the MSE of the speech distortion, i.e., $J_{\mathrm{sd}}\left[\mathbf{h}(k,m)\right]$, under the constraint that the MSE between the filtered echo and the target residual echo, i.e., $J_{\mathrm{re}}\left[\mathbf{h}(k,m)\right]$, is kept below some given threshold $\varepsilon \ge 0$. Mathematically, the optimal filter is expressed as

$$\begin{aligned}\mathbf{h}_{\mathrm{o}}(k,m) = \underset{\mathbf{h}(k,m)}{\operatorname{argmin}}\ & J_{\mathrm{sd}}\left[\mathbf{h}(k,m)\right] \\ \text{subject to}\ & J_{\mathrm{re}}\left[\mathbf{h}(k,m)\right] \le \varepsilon.\end{aligned} \tag{16}$$

Solving the above problem using the Lagrange method with a multiplier μ yields the parametric Wiener filter:

$$\mathbf{h}_{\mathrm{PWF}}(k,m) = [\mathbf{\Phi_u}(k,m) + \mu\mathbf{\Phi_y}(k,m)]^{-1} \times [\mathbf{\Phi_u}(k,m)\mathbf{i}_1 + \mu\alpha\mathbf{\Phi_y}(k,m)\mathbf{i}_1]. \quad (17)$$

This parametric Wiener filter depends only on the second-order statistics of the subband signals of the desired near-end speech and the echo.

3.2. Properties of the Parametric Wiener Filter

The filter given in (17) can be decomposed into a weighted sum of two Wiener filters: one extracts the near-end speech and the other one extracts the echo signal. By defining the modified correlation matrix of the microphone signal as $\widetilde{\mathbf{\Phi}}_{\mathbf{d}}(k,m) = \mathbf{\Phi_u}(k,m) + \mu\mathbf{\Phi_y}(k,m)$, we can rewrite (17) as

$$\begin{aligned}\mathbf{h}_{\mathrm{PWF}}(k,m) &= \widetilde{\mathbf{\Phi}}_{\mathbf{d}}^{-1}(k,m)\mathbf{\Phi_u}(k,m)\mathbf{i}_1 + \\ &\quad \alpha\widetilde{\mathbf{\Phi}}_{\mathbf{d}}^{-1}(k,m)\mu\mathbf{\Phi_y}(k,m)\mathbf{i}_1 \\ &= (1-\alpha)\,\mathbf{h}_{\mathrm{T},\mu}(k,m) + \alpha\mathbf{i}_1. \end{aligned} \quad (18)$$

where $\mathbf{h}_{\mathrm{T},\mu}(k,m) = \widetilde{\mathbf{\Phi}}_{\mathbf{d}}^{-1}(k,m)\mathbf{\Phi_u}(k,m)\mathbf{i}_1$ is a tradeoff Wiener filter considering the interframe correlation, which can be derived similarly as for the multichannel case in [20]. From (18), it is clear that by introducing the residual echo control parameter α, the obtained filter can be seen as a weighted sum of a tradeoff Wiener and the unit filters.

If $\alpha = 0$, the parametric Wiener filter $\mathbf{h}_{\mathrm{PWF}}(k,m)$ degenerates to the well-known tradeoff Wiener filter [20], [21], where the target signal is the desired near-end speech only. The Lagrangian multiplier μ adjusts the quality of the near-end speech, i.e., it enables control of the level of over- or underestimation. For $\mu = 1$, by adjusting the value of α from 0 to 1, the parametric Wiener filter can control the level of the residual echo, which is similar to the filter derived in [22], [23] in the context of binaural hearing aids. For both $\mu = 1$ and $\alpha = 0$, the parametric Wiener filter degenerates to the conventional Wiener filter that considers the correlation between L consecutive time frames. In this case and if $L = 1$, the parametric Wiener filter degenerates to a simple Wiener gain filter as in [9]–[16].

4. EXPERIMENTS

In this section, we study the developed parametric Wiener filter through experiments and investigate the impact of the value of the parameters μ and α in (17) on the echo suppression performance.

4.1. Experimental Setup and Estimation of the Statistics

The far-end and near-end speech signals used in the experiments are recorded in an anechoic room from male and female talkers separately. They are sampled at 8 kHz and quantized with 16 bits. The echo signal is generated by convolving the far-end speech with an impulse response measured in a room with a reverberation time T_{60} of approximately 380 ms and the distance between the loudspeaker and microphone is 1.5 m. The microphone signal is then synthesized by superimposing this echo with the near-end signal and white Gaussian noise is then added to control the SNR to be 30 dB.

The overlap-add technique is employed in the implementation with a DFT length of $K = 256$ and 75% overlap between neighboring frames. To overcome the aliasing problem, a Kaiser window is applied both before the STFT and after the inverse STFT.

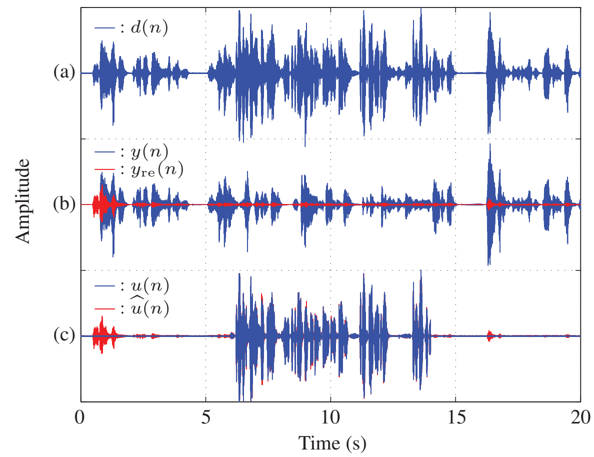

Fig. 2. An example for different signals processed by a conventional Wiener filter: (a) microphone signal $d(n)$, (b) echo $y(n)$ and residual echo $y_{\mathrm{re}}(n)$, and (c) near-end signal $u(n)$ and estimated near-end signal $\widehat{u}(n)$. $\mu = 1$, $\alpha = 0$, $\lambda = 0.35$, and $L = 4$.

To implement the parametric Wiener filter in (17) for echo suppression, the paramount issue is to have a good estimate of the correlation matrices $\mathbf{\Phi_u}(k,m)$ and $\mathbf{\Phi_y}(k,m)$. In all experiments, an NPVSS NLMS algorithm is used to estimate the echo component $\widehat{Y}(k,m)$ in every frequency bin k and time frame m in the STFT domain [4], [24]. The $\mathbf{\Phi_d}(k,m)$ and $\mathbf{\Phi}_{\widehat{\mathbf{y}}}(k,m)$ matrices are then computed with a rank-1 update approach [25]:

$$\mathbf{\Phi_z}(k,m) = \lambda\mathbf{\Phi_z}(k,m-1) + (1-\lambda)\mathbf{z}(k,m)\mathbf{z}^H(k,m), \quad (19)$$

where $\lambda \in (0,1)$ is a forgetting factor, and $\mathbf{z}(k,m) \in \{\mathbf{d}(k,m), \widehat{\mathbf{y}}(k,m)\}$. Note that, in our simulations, we use the first 100 frames to compute the initial estimates of the $\mathbf{\Phi_d}(k,m)$ and $\mathbf{\Phi}_{\widehat{\mathbf{y}}}(k,m)$ matrices with a short-time average. Finally, $\mathbf{\Phi_u}(k,m)$ is estimated as $\mathbf{\Phi_u}(k,m) = \mathbf{\Phi_d}(k,m) - \mathbf{\Phi}_{\widehat{\mathbf{y}}}(k,m)$, which is ensured to be positive semi-definite. For numerical stability, we invert the correlation matrices by a truncated eigenvalue decomposition (EVD) as in [26]. We test the performance of the parametric Wiener filter during single- and double-talk (between 6 and 14 seconds). While it is a very important issue, how to effectively detect the double-talk situation is beyond the scope of this paper. So, we set aside the double-talk detection issue and directly use an "ideal" detector to stop the filter adaptation during double-talk periods.

In all experiments, we set $\lambda = 0.35$ and $L = 4$. To visualize the echo suppression performance, Fig. 2 plots the residual echo and estimated near-end signals obtained with the parametric Wiener filter, i.e., (17) with $\mu = 1$ and $\alpha = 0$. The microphone, echo, and near-end signals are also plotted for comparison. As can be seen from Fig. 2, the parametric Wiener filter achieves a significant amount of echo suppression while maintaining the near-end signal after processing close to the original one.

4.2. Performance Measures

To evaluate the echo suppression performance of the parametric Wiener filter, we use three performance metrics: echo suppression gain (ESG), near-end speech distortion index, and the perceptual evaluation of speech quality (PESQ) measure [27]. The ESG, which is also called echo-return loss enhancement (ERLE) in AEC [3], [28], is defined as

$$\mathrm{ESG}(n) = 10\log_{10}\frac{\mathcal{E}\{y^2(n)\}}{\mathcal{E}\{y_{\mathrm{re}}^2(n)\}}, \quad (20)$$

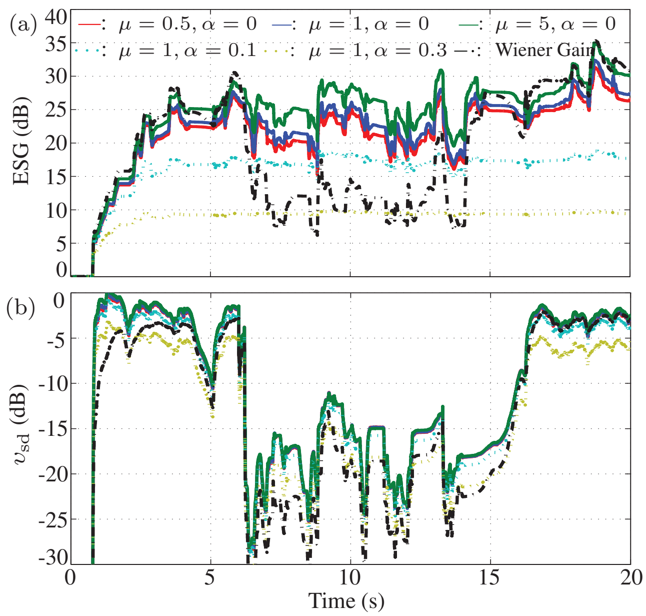

Fig. 3. Echo suppression performance of the parametric Wiener filter: (a) ESG and (b) speech distortion index, υ_{sd}. $\lambda = 0.35$ and $L = 4$. The performance of the traditional Wiener gain filter (with $\mu = 1$, $\alpha = 0$, and $L = 1$) is also plotted for comparison.

and the near-end speech distortion index [17], [29] is given by

$$\upsilon_{\mathrm{sd}}(n) = 10 \log_{10} \frac{\mathcal{E}\left\{[u(n) - u_{\mathrm{f}}(n)]^2\right\}}{\mathcal{E}\left\{u^2(n)\right\}}, \tag{21}$$

where $y_{\mathrm{re}}(n)$ and $u_{\mathrm{f}}(n)$ are the time-domain signals reconstructed from $Y_{\mathrm{re}}(k, m)$ in (8) and $U_{\mathrm{f}}(k, m)$ in (7), respectively.

4.3. Echo Suppression Performance as a Function of μ and α

The first experiment investigates the influence of the value of the parameters μ and α in (18) on the echo suppression performance. The results as a function of time for different values of μ, are presented in Fig. 3. For comparison, the experimental results of the traditional Wiener gain filter (with $\mu = 1$, $\alpha = 0$, and $L = 1$, as mentioned in Section 3.2) are also plotted.

First, we set $\alpha = 0$, vary the value of μ from 0.5 to 5, and the results are plotted in solid lines in Fig. 3. As can be seen, a larger value of μ leads to a higher ESG; but the speech distortion index increases as well. So, the parameter μ allows a compromise between echo attenuation and near-end speech distortion, i.e., by choosing a larger value of μ, it is possible to achieve a higher ESG by slightly increasing the speech distortion, and vice versa.

Then, we set μ to 1, adjust the value of α from 0 to 0.3, and the results are also plotted in Fig. 3 (with dotted lines). It is clearly seen that both the ESG and speech distortion index decrease as the value of α increases.

For a fixed value of μ and α, e.g., $\mu = 1$ and $\alpha = 0$, one can see that the PWF can achieve more than 25 dB echo attenuation while the near-end speech distortion is less than -20 dB during the double-talk situation, which is sufficient for many applications.

Figure 3 also plots the results for the traditional Wiener gain filter, i.e., $L = 1$. Comparing the results of the parameter Wiener filter with $L = 4$ with those of the traditional Wiener gain, one can see that the PWF yields a much higher ESG than the traditional Wiener gain (the dark line) during double-talk at the cost of adding slightly more speech distortion to the desired near-end speech. This shows the advantage of using the interframe correlation in echo suppression.

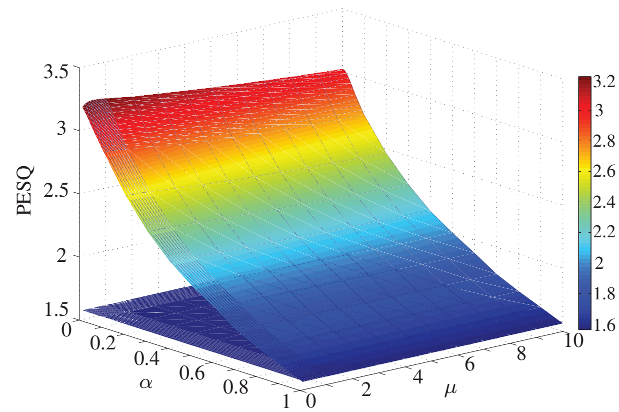

Fig. 4. The PESQ score of the parametric Wiener filter as a function of μ and α with $\lambda = 0.35$ and $L = 4$. The higher colorful surface is the PESQ score calculated between $u(n)$ and $\widehat{u}(n)$, whereas the blue flat surface is the PESQ score calculated between $u(n)$ and $d(n)$.

4.4. Speech Quality Evaluation with PESQ

In the third experiment, we evaluate the quality of the estimated near-end speech through the PESQ measure as a function of the parameter μ and α. The PESQ score calculated between the near-end and microphone signals are used as a benchmark for comparison. The results are plotted in Fig. 4. One can see that, for a small value of α ($\alpha \leq 0.05$), the PESQ score is not a monotonic function with respect to μ. It first increases and then decreases as the value of μ increases. The underlying reason can be explained as follows. As shown in Section 4.3, both ESG and the near-end speech distortion increases with the value of μ. When μ is smaller, increasing its value can help increase ESG, thereby improving the PESQ score. But when the value of μ is large, the near-end speech distortion is also high, which causes degradation in PESQ.

For all values of μ, if $\alpha = 1$, the filter in (18) becomes a unit one, so the echo signal is not attenuated and remains in the target signal. In this case, the filtering processing does not change the PESQ score as confirmed by Fig. 4. As the value of α increases from 0 to 1, the amount of echo attenuation decreases, so does the PESQ score, which is clearly seen in Fig. 4.

The highest PESQ score is obtained when μ is approximately 0.5 and $\alpha = 0$ in all the studied experimental conditions. This answers the question why we are interested in the parametric Wiener filter approach. Briefly, the parametric Wiener filter can yield better echo suppression performance as compared to the traditional Wiener filter if the value of the parameters μ and α are properly chosen.

5. CONCLUSIONS

This paper investigated the problem of AES in the STFT domain. Based on the use of interframe correlation, an AES framework was formed and a parametric Wiener filter was then derived by minimizing the near-end speech distortion with a constraint on the level of the residual echo. By adjusting the two parameters μ and α in the parametric Wiener filter, one can control the level of the residual echo and identify a proper tradeoff between the amount of echo attenuation and the degree of near-end speech distortion for better perceived speech quality. Simulation results verified the properties of the developed parametric Wiener filter for AES.

6. REFERENCES

[1] C. Breining, P. Dreiscitel, E. Gänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control–an application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, pp. 42–69, July 1999.

[2] J. Benesty and Y. Huang, *Adaptive signal processing: applications to real-world problems*. Berlin, Germany: Springer-Verlag, 2003.

[3] C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*. San Rafael: Morgan & Calypool, 2010.

[4] E. A. P. Habets, S. Gannot, and I. Cohen, "Robust early echo cancellation and late echo suppression in the STFT domain," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 14–17, 2008.

[5] G. W. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 67–70, 2003.

[6] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[7] A. Schwarz, C. Hofmann, and W. Kellermann, "Spectral feature-based nonlinear residual echo suppression," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–4, 2013.

[8] Y. Guelou, A. Benamar, and P. Scalari, "Analysis of two structures for combined acoustic echo cancellation and noise reduction," in *Proc. Euro. Signal Process. Conf.*, pp. 1–4, 1996.

[9] C. Avendano, "Acoustic echo suppression in the STFT domain," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 175–178, 2001.

[10] F. Wallin and C. Faller, "Perceptual quality of hybrid echo canceler/suppressor," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 157–160, 2004.

[11] C. Faller and J. Chen, "Suppressing acoustic echo in a spectral envelope space," *IEEE Trans. Speech, Audio Process.*, vol. 13, pp. 1048–1062, Sep. 2005.

[12] C. Faller and C. Tournery, "Estimating the delay and coloration effect of the acoustic echo path for low complexity echo suppression," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 53–56, 2005.

[13] C. Faller and C. Tournery, "Robust acoustic echo control using a simple echo path model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 281–284, 2006.

[14] A. Favrot, C. Faller, M. Kallinger, and M. Schmidt, "Acoustic echo control based on temporal fluctuations of short-time spectra," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 1–4, 2008.

[15] A. Favrot, C. Faller, and F. Kuech, "Modeling late reverberation in acoustic echo suppression," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 1–4, 2012.

[16] Y. S. Park and J. H. Chang, "Frequency domain acoustic echo suppression based on soft decision," *IEEE Signal Process. Lett.*, vol. 16, pp. 53–56, Jan. 2009.

[17] H. Huang, J. Benesty, J. Chen, K. Helwani, and H. Buchner, "A study of the MVDR filter for acoustic echo suppression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 615–619, 2013.

[18] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, pp. 113–120, Apr. 1979.

[19] K. Helwani, H. Buchner, J. Benesty, and J. Chen, "A single-channel MVDR filter for acoustic echo suppression," *IEEE Signal Process. Lett.*, vol. 20, pp. 351–354, Apr. 2013.

[20] J. Benesty, J. Chen, and E. Habets, *Speech Enhancement in the STFT Domain*. Berlin, Germany: Springer-Verlag, 2011.

[21] W. Zhang, J. Chen, and J. Benesty, "A reduced-rank approach to single-channel noise reduction," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 293–297, 2014.

[22] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, pp. 1579–1585, Apr. 2007.

[23] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *J. Acoust. Soc. Am.*, vol. 125, pp. 360–371, Jan. 2009.

[24] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, pp. 581–684, Oct. 2006.

[25] K. Yu, "Recursive updating the eigenvalue decomposition of a covariance matrix," *IEEE Trans. Signal Process.*, vol. 39, pp. 1136–1145, May 1991.

[26] H. Huang, L. Zhao, J. Chen, and J. Benesty, "A minimum variance distortionless response filter based on the bifrequency spectrum for single-channel noise reduction," *Digit. Signal Process.*, vol. 33, pp. 169–179, Oct. 2014.

[27] ITU-T P.862, "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*, 2001.

[28] C. Huemmer, C. Hofmann, R. Maas, A. Schwarz, and W. Kellermann, "The elitist particle filter based on evolutionary strategies as novel approach for nonlinear acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 1315–1319, 2014.

[29] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction wiener filter," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, pp. 1218–1234, July 2006.