# An Improved Low-Complexity Echo Suppression Algorithm Based on the Acoustic Coloration Effect

Ruxue Guo, Tao Jiang, Qingyun Wang, Ruiyu Liang, and Cairong Zou

***Abstract*—The echo cancellation algorithm plays a vital role in the acoustic systems, but it has the defects of long convergence time and high complexity. In this study, a low-complexity echo suppression scheme based on the coloration effect of the feedback path is introduced to address these challenges. The acoustic feedback path is modeled as a coloration filter, and the spectrum of the echo signal can be estimated according to the coefficients of the coloration filter and the spectrum of the loudspeaker signal. This method has two main innovations. Firstly, a parametric Wiener filter based on the prior signal-to-echo ratio (SER) is employed to reduce the gain fluctuation and the residual echo. Secondly, we apply a parameter adaptation strategy based on posterior SER and a correlation factor to improve the echo suppression capability and the speech quality. Experimental results show that the ERLE in the proposed algorithm is improved by 15dB compared with that in the classical frequency-domain adaptive filtering (FDAF) algorithm. The convergence time is shortened from 0.5 s to 0.1 s, and the computational complexity decreases significantly. Also, this scheme simulating the acoustic path with a coloration filter has better robustness for acoustic path mutation.**

***Index Terms*—acoustic echo cancellation, acoustic echo suppression, acoustic coloration effect, the parametric Wiener filter**

## I. INTRODUCTION

ACOUSTIC echo cancellation is a built-in module for many acoustic devices such as conference phones, hearing aids, and smart speakers[1], [2]. The conventional methods usually apply an adaptive filter to estimate the acoustic path of a loudspeaker-enclosure-microphone system. These acoustic echo cancellation algorithms using adaptive filters mainly involve the normalized least mean square (NLMS), the affinity propagation (AP), the recursive least square (RLS), and their improved algorithms[3], [4], [5], [6]. The schemes based on adaptive filters take a long time and large calculation amount to converge from zero to a stable acoustic path. Thus, the residual echo seriously affects the user experience during the initial convergence of the system.

In recent years, acoustic echo cancellation (AEC) algorithms based on neural networks have become a research focus[7], [8], [9]. Mehdi Bekrani designed a linear single-layer feedforward neural network to de-correlate the input signal and reference signal and to achieve a fast convergence rate and low misalignment[7]. Seo suggested a stacked deep neural network (DNN) frame, which includes a set of DNNs for noise suppression (NS) and AES[8]. The DNN is trained to map noisy speech signals to clear speech signals in the scheme. Then, the mapped speech is input into the DNN, which is designed for AES to suppress the echo signal. Hao Zhang proposed a bidirectional recurrent long short-term memory network (BiLSTM) to separate target speech from echoes[9]. The algorithms mentioned above can obtain better performance than traditional algorithms due to a large amount of training. However, the high computational complexity of neural networks makes it difficult to implement them on low-power acoustic devices. Therefore, the acoustic echo suppression (AES) algorithm based on spectral correction has been widely explored by researchers, with the advantages of ultra-low computational complexity and fast convergence characteristics[10], [11], [12].

Similar to noise suppression based on spectral subtraction [13], [14] and traditional post-nonlinear filter[15], [16], the AES system corrects the signal spectrum using a series of parameters in order to achieve echo suppression in the frequency domain[17]. The echo component is caused by feedback from a loudspeaker signal to a microphone along the acoustic path. When the echo signal reaches the microphone, the loudspeaker signal is time-shifted and the amplitude spectrum is modulated. This phenomenon is called the coloration effect, which is caused by early sound reflection in the room. Late reflection is generally ignored in this model[18]. A perceptual acoustic echo suppression (PAES)[19] was proposed by Wallin to estimate the spectral envelope of the echo signal according to the frequency selection characteristics of the human auditory system. The Wiener filter was used to suppress the echo spectrum[10], [20]. Ying proposed a new AES scheme based on the beta mixture model (BMM)[11], which estimated the probability of near-end speech existence under Bayesian rule and introduced the probability of speech existence on the basis of Wiener filtering in the scheme. In comparison with the echo suppression schemes based on the AEC or deep neural network, the main advantage of the AES algorithm with a spectrum correction technique is that the acoustic feedback can be suppressed at low computational complexity. At the same time, the algorithm also has the characteristics of fast convergence and anti-path mutation, and therefore, it can be widely applied in low-power acoustic systems. However, the AES may cause speech distortion and sound-quality degradation. Achieving the optimal tradeoff between echo suppression and speech distortion is still a difficult problem.

To compromise the echo attenuation and speech distor-

Manuscript received November 26, 2021; revised May 13, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant No. 2020YFC2004003 and 2020YFC2004002, the National Natural Science Foundation of China under Grant No. 62001215, the Plan Project of Changzhou Science and Technology (social development) under Grant No.CE20195036.

Ruxue Guo is a doctoral candidate of the School of Information Science and Engineering, Southeast University, China (email: grx0904@sina.com)

Tao Jiang is a graduate student of the School of Information Science and Engineering, Southeast University, China (email: 18351966578@163.com)

Qingyun Wang is a Professor of Nanjing Institute of Technology, Nanjing, China (e-mail: wangqingyun@njit.edu.cn)

Ruiyu Liang is a Professor of Nanjing Institute of Technology, Nanjing, China (e-mail: liangry@njit.edu.cn)

Cairong Zou is a Professor of the School of Information Science and Engineering, Southeast University, China (e-mail: cairong@seu.edu.cn)

tion of the AES system, we proposed an improved low-complexity real-time echo suppression algorithm based on the acoustic coloration effects in this paper. A parametric Wiener filter based on prior SER estimation was employed to conduct the real-time echo suppression, and the acoustic echo spectrum was obtained by a coloration filter. Furthermore, a parameter adaptation strategy based on posterior SER and double-talk decision factors was integrated to update the parameters of the gain control function of the parametric Wiener filter. The simulation results indicated that the proposed parameter adaptive method effectively improved the echo suppression performance. The main contributions of this work are three-fold, as follows:

- We applied the parametric Wiener filter to suppress feedback from the acoustic system, and the coloration effect was considered when estimating the echo spectrum.
- We introduced a novel parameter adaptation strategy to adjust the echo attenuation. The parameters were updated according to posterior SER and a frequency-domain double-talk decision factor. The factor was designed to determine the speech interaction state.
- The experimental results illustrated that a 16 dB ERLE improvement was achieved in comparison with the classical FDAF algorithm, and a 4 dB ERLE improvement was achieved in comparison with the Wiener filtering algorithm based on prior SER while maintaining fast initial convergence characteristics and robustness of anti-path mutation.

## II. Wiener Filtering Echo Suppression Algorithm Based on Coloration Effect

### A. Echo suppression algorithm based on Wiener filtering

The echo spectrum is estimated by the echo suppression algorithm based on spectral correction to eliminate the acoustic feedback in combination with the suppression function, which differs from the algorithm based on an adaptive filter and no longer requires a high-order filter. The microphone pick-up signal $Y(k)$ is given by

$$Y(k) = S(k) + D(k) \tag{1}$$

where $S(k)$ and $D(k)$ are frequency-domain expressions of subscriber-side speech and echo, respectively, and $k$ is the frequency bin index. In addition, the estimation of the near-end speech is expressed as

$$\hat{S}(k) = G_{WN}(k)Y(k) \tag{2}$$

where $G_{WN}(k)$ is the gain function of the Wiener filter and is defined as:

$$G_{WN}(k) = \frac{|Y(k)|^2 - |\hat{D}(k)|^2}{|Y(k)|^2} = \frac{\lambda_s(k)}{\lambda_s(k) + \lambda_d(k)} \tag{3}$$

where $|\hat{D}(k)|$ is the estimated echo amplitude spectrum; $|Y(k)|$ is the amplitude spectrum of the microphone pick-up signal; and $\lambda_s(k)$ and $\lambda_d(k)$ are the power spectra of the near-end speech and echo, respectively. Prior SER is introduced to the Wiener filter algorithm to avoid residual echo such as music noise. The prior SER $\zeta(k)$ and posterior SER $\gamma(k)$ are defined as

$$\zeta(k) = \frac{E[|S(k)|^2]}{\lambda_d(k)} \tag{4}$$

$$\gamma(k) = \frac{|Y(k)|^2}{\lambda_d(k)} \tag{5}$$

'Prior' and 'posterior' refer to the information that depends on the previous and current frame, respectively. The parametric Wiener filter in the frequency domain can be defined to replace $G_{WN}(k)$[21].

$$H(k) = \left(\frac{\lambda_s(k)}{\lambda_s(k) + \alpha\lambda_d(k)}\right)^\beta \tag{6}$$

where $\alpha$ and $\beta$ are the filter parameters. If $\alpha = \beta = 1$, Equation (6) degenerates to the traditional Wiener filter; if $\alpha = 1$ and $\beta = 1/2$, it turns into the square root Wiener filter. Hence, parameters $\alpha$ and $\beta$ directly affect the attenuation characteristics of the Wiener filter. Bringing Equation (4) into (6), the Wiener filter based on prior SER is given by

$$H(k) = \left(\frac{\zeta(k)}{\zeta(k) + \alpha}\right)^\beta \tag{7}$$

Then, the smoothing factor $\eta_1$ is introduced to estimate the prior SER. The estimation of the prior SER $\zeta_i(k)$ is generally described as

$$\hat{\zeta}_i(k) = \eta_1\zeta_{i-1}(k) + (1 - \eta_1)\varphi(\gamma_i(k) - 1) \tag{8}$$

where $i$ represents the frame number. In addition, $\varphi(u) = u$, if $u > 0$; otherwise, $\varphi(u) = 0$. The prior SER $\zeta_i(k)$ of $i$ frame in Equation (7) can be estimated by the weighted combination of prior SER $\zeta_{i-1}(k)$ of the $i-1$ frame and the posterior SER of the $i$ frame. In addition, the selection of the smoothing factor $\eta_1$ has a great impact on the final performance. When $\eta_1$ is close to 1, the gain function is smoother and the residual signal is lower. However, the subscribe-side voice quality is seriously degraded during double-talk. Empirically, $\eta_1$ can be set to 0.6[22]. Therefore, the subscribe-side speech spectrum based on the prior SER can be obtained as

$$|\hat{S}_i(k)| = \hat{H}_i(k)|Y(k)| = \frac{\hat{\zeta}_i(k)}{\hat{\zeta}_i(k) + 1}|Y(k)| \tag{9}$$

### B. Echo spectrum estimation based on coloration effect

In audio processing, 'white' refers to equal energy of each frequency bin, and 'coloration' generally means that some frequencies are attenuated, whereas other frequencies are amplified or not attenuated. Therefore, the acoustic echo path can be regarded as a coloration filter in the AES system and the echo signal can be estimated when the response and delay parameters of the coloration filter are given[23]. An improved spectrum estimation method based on the coloration effect is introduced in this paper. The echo is obtained by estimating the coefficients of the coloration filter, which enhances the robustness in comparison with the conventional AEC scheme of estimating the acoustic echo path.

In the AES system, the acoustic echo path is divided into three parts: 1) direct path (direct transmission from loudspeaker to microphone), 2) early reflection path, and 3) high-density late reflection. The coloration effect of the audio signal is introduced by direct sound and early reflection. Late reflection does not or hardly colors the signal. Therefore, it is reasonable to include only the direct sound and early reflection parts to estimate the echo signal effectively in the algorithm. A coloration filter is defined to simulate the acoustic echo path as follows.

$$|\hat{D}_i(k)| = \hat{G}_i(k)|X_i(k)| \tag{10}$$

where $\hat{G}_i(k)$ are the coefficients of the coloration filter, which represent the spectral coloration effect of an acoustic echo path on a loudspeaker signal spectrum.

The expression of least square estimation of the coloration filter is

$$G_i(k)=\left|\frac{E\{X_i^*(k)Y_i(k)\}}{E\{X_i^*(k)X_i(k)\}}\right| \tag{11}$$

where $X_i^*(k)$ is the complex conjugate of $X_i(k)$. As the acoustic echo is a time-varying signal, it can be calculated iteratively as

$$\hat{G}_i(k)=\frac{C_i(k)}{R_i(k)} \tag{12}$$

where

$$C_i(k) = \alpha_c C_{i-1}(k) + (1-\alpha_c)\,|X_i^*(k)Y_i(k)| \tag{13}$$

$$R_i(k) = \alpha_R R_{i-1}(k) + (1-\alpha_R)\,|X_i^*(k)X_i(k)| \tag{14}$$

where $\alpha_c$ and $\alpha_R$ are adjustable smoothing factors.

## III. Improved Parametric Wiener Filter Echo Suppression Algorithm

To improve the performance of the feedback suppression algorithm, we proposed a parametric Wiener filter, which is detailed in this section. In the traditional Wiener filter and the Wiener filter based on prior SER, the parameters $\alpha$ and $\beta$ of Equation (7) are fixed. However, there are usually double-talk and single far-end statuses in the actual AES system. In a single far-end state, the microphone picks only echoes from the loudspeaker by a far-end source, and the SER is generally less than 0 dB. It is necessary to eliminate the echoes as far as possible. In a double-talk state, the microphone simultaneously receives a near-end signal and an echo signal. Sometimes, the energy of the echo signal is stronger than that of the near-end signal. In this situation, the attenuation of the AES system should not be very large, even if SER is less than 0 dB. Otherwise, the near-end speech is seriously distorted or truncated. Therefore, the parameter of the Wiener filter should be adjusted adaptively according to the interaction state of the speech.

A parameter adaptation strategy based on posterior SER is introduced to meet the requirement mentioned above, and a double-talk decision factor is fused into the adaptation process to improve the performance. The block diagram of the proposed algorithm is shown in Fig. 1.

### A. Parameter adaptation based on posterior SER

Compared with $\beta$, the parameter $\alpha$ is more flexible and effective in the gain adjustment of a parametric Wiener filter. Thus, the $\alpha$ is designed to be an adaptive factor $\alpha_i$, which is adjusted by the SER of the acoustic environment.

$$\alpha_i = f_\alpha(\gamma_i(k)) \tag{15}$$

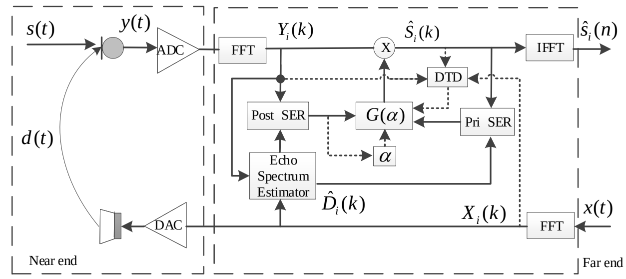


Fig. 1. The block diagram of the improved parametric Wiener filtering echo suppression algorithm

where $\gamma_i(k)$ is the posterior SER of the $i$ frame, and $i$ is the frame number. A smoothing factor $\alpha_\gamma$ is set to prevent excessive fluctuations in the gain in the parametric Wiener filter.

$$\tilde{\gamma}_i(k) = \alpha_\gamma \cdot \tilde{\gamma}_{i-1}(k) + (1-\alpha_\gamma)\cdot 10\log\gamma_i(k) \tag{16}$$

In general, $\alpha_\gamma$ uptakes the value from 0.992 to 0.998. $\tilde{\gamma}_i(k)$ is the posterior SER of the $i$ frame after smoothing, and the SER is converted to the decibel domain. The function $f_\alpha(\cdot)$ represents a non-linear function that is used to calculate the parameter $\alpha_i$ on the basis of the posterior SER. $\alpha_i$ is designed as a piecewise function to reduce the calculation complexity, as follows.

$$a_i = \begin{cases} 5, \tilde{\gamma}_i(k) < -5dB \\ 4-\dfrac{1}{5}\tilde{\gamma}_i(k), -5dB \le \tilde{\gamma}_i(k) \le 15dB \\ 1, \tilde{\gamma}_i(k) > 15dB \end{cases} \tag{17}$$

The reasons why this function $f_\alpha(\cdot)$ was chosen are as follows: 1) when posterior SER is relatively small, such as in the far-end state, or when the energy of the echo signal is far stronger than that of the near-end signal, the algorithm requires a large parametric $\alpha$ to suppress echoes; 2) when the posterior SER is large, such as in a double-talk state or a single near-end state, the algorithm requires a smaller parameter $\alpha$ to avoid damage to the near-end signal and to ensure the quality of speech.

### B. A frequency-domain double-talk decision factor

In a double-talk state, the near-end speech has a masking effect on the residual echo. In this case, the fluency of near-end speech is important and the existence of residual echo is allowed. Therefore, the suppression curve of the parametric Wiener filter should be set to moderate attenuation in this case. A frequency-domain double-talk detector is proposed in this section, which designs a correlation decision factor and applies it to the adaptation process of the parametric Wiener filter.

The double-talk detection includes two cross-correlation variables: 1) the correlation coefficient $\rho_{yd}(k)$ between the microphone signal and the estimated echo signal, 2) the correlation coefficient $\rho_{ys}(k)$ between the microphone signal and the output signal of the echo suppressor.

$$\rho_{yd}(k) = \frac{\lambda_{yd}(k)}{\sqrt{\lambda_y(k)}\cdot\sqrt{\lambda_d(k)}} \tag{18}$$

$$\rho_{ys}(k) = \frac{\lambda_{ys}(k)}{\sqrt{\lambda_y(k)}\cdot\sqrt{\lambda_s(k)}} \tag{19}$$

where $\lambda_y(k)$ is the power spectral density (PSD) of the microphone input signal, $\lambda_d(k)$ is the PSD of the estimated echo signal, and $\lambda_s(k)$ is the PSD of the echo suppressor output signal. According to the orthogonality principle, the estimated echo signal is close to the true echo when the filter completely simulates the acoustic feedback path. For the single-far end, the correlation coefficient of the microphone input signal $Y(k)$ and the estimated echo signal $\hat{D}(k)$ is close to 1. Therefore, $\rho_{yd}(k)$ tends to be 1. On the other hand, the output signal $\hat{S}(k)$ tends to be 0, and the microphone input signal is uncorrelated to $\hat{S}(k)$. Therefore, $\rho_{ys}(k)$ between the echo suppressor output $\hat{S}(k)$ and the input signal of the microphone tends to be 0. In contrast, $\rho_{yd}(k)$ decreases and $\rho_{ys}(k)$ increases in the double-talk state.

In order to make the changes in $\rho_{yd}(k)$ and $\rho_{ys}(k)$ be in the same direction, we define

$$\tilde{\rho}_{ys}(k) = 1 - \rho_{ys}(k) \tag{20}$$

Both $\rho_{yd}(k)$ and $\tilde{\rho}_{ys}(k)$ decrease under double-talk. In addition, when we combine correlation coefficients $\rho_{yd}(k)$ and $\tilde{\rho}_{ys}(k)$ by weighting factor $\alpha_{\varpi}$, the correlation decision factor $\varpi(k)$ is given by

$$\varpi(k) = \alpha_{\varpi}\rho_{yd}(k) + (1-\alpha_{\varpi})\tilde{\rho}_{ys}(k) \tag{21}$$

Subsequently, the correlation decision factor $\varpi(k)$ is incorporated into the parametric Wiener filter:

$$H(k) = \left(\frac{\lambda_s(k)}{\lambda_s(k)+\alpha\lambda_d(k)}\right)^{\max\{a\cdot\varpi(k),\beta_{\min}\}} \tag{22}$$

where $a$ is a magnification factor that is generally set to 2. $\beta_{\min}$ is the minimum value of $\beta$. For a single far-end state without the near-end speech, $\rho_{yd}(k)$ and $\tilde{\rho}_{ys}(k)$ both tend towards 1, and $a\cdot\varpi(k)$ tends towards 2. The attenuation of the parametric Wiener filter becomes larger. For a double-talk state, $\rho_{yd}(k)$ and $\tilde{\rho}_{ys}(k)$ both tend towards 0. Thus, $a\cdot\varpi(k)$ tends towards 0, and the attenuation of the parametric Wiener filter becomes weak to protect the near-end speech from excessive suppression. $\beta_{\min}$ is a constant that is set to $[0.5 \quad 1.0]$. When $\beta_{\min} = 0.5$, the parametric Wiener filter changes to the square root filter. When operating on double-talk, the suppression is low, and the quality of the near-end speech is guaranteed. When $\beta_{\min} = 1$, the parametric Wiener filter degenerates to the basic Wiener filter.

### C. Calculation complexity

Computational complexity is an important factor for low-power acoustic processors. In general, fast Fourier transform (FFT) is the most computationally expensive part for frequency-domain algorithms. For the AEC scheme based on frequency domain adaptive filter (FLMS or PBFDAF), an iteration involves at least two FFTs and two IFFTs. Each FFT or IFFT requires $L_w\log_2 L_w$ real multiplications, and the $L_w$ is the filter length. In contrast, the scheme based on spectral correction (AES) needs three Fourier forward/inverse transforms and requires $3N_{f1}\log_2 N_{f1}$ real multiplications, and the $N_{f1}$ is the frame length. The acoustic feedback path from speaker to microphone in the system is, in general, multiple times the length of the frame length. For example, the length of the adaptive filter takes $L_w = 512$, and the frame length $N_{f1} = 128$. Then, $L_w\log_2 L_w$ requires 4608 real multiplications, and $N_{f1}\log_2 N_{f1}$ only requires 896 real multiplications. Therefore, the computational complexity of the proposed algorithm is much smaller than the algorithms based on the frequency domain adaptive filter.

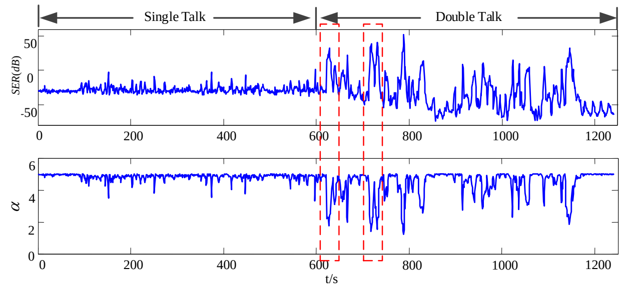


Fig. 2. Real-time values of $\alpha$ and SER in single-far and double-talk state

## IV. Experimental Results and Analysis

### A. Experimental configuration

The tested acoustic echo path was generated by the room pulse generation toolbox. The length, width, and height of the room were set to 5, 4, and 3 m, respectively. The acoustic echo path was truncated to 2048 orders. The test speeches were selected from the listening materials of the Mandarin Proficiency Test[24]. The sampling rate of the test signal was 16 kHz with 16-bit accuracy, and the duration of each file was approximately 20 s.

The proposed algorithm in this paper was compared with the traditional FDAF method, basic Wiener filtering method, and Wiener filtering method based on prior SER. The processed frame length of the algorithm was 128 samples, with a frameshift of 50%. The length of the adaptive filter in FDAF was 2048 orders. The echo return loss enhancement (ERLE) was calculated to evaluate the performance.

$$ERLE = 10\lg\frac{d^2(n)}{e^2(n)} \tag{23}$$

where $e(n)$ is the error signal, and $d(n)$ is the desired signal. A larger ERLE indicates a better echo cancellation performance.

### B. Parameter $\alpha$ varies according to the SER

Here we randomly choose a speech signal as an example to illustrate how parameter $\alpha$ changes according to SER. Fig. 2 shows the performances of SER and $\alpha$ in the single-far state and the double-talk state. In the single-far state, SER was low(around -40 dB), and $\alpha$ was close to the maximum; therefore, the attenuation of the parameter Wiener filter was the largest. In the double-talk state, SER increased and was followed by near-end voice energy.

When near-end voice energy was large and echo signal was weak, SER was high, generally larger than 0 dB. Inversely, $\alpha$ decreased and the attenuation of the parameter Wiener filter was weakened to avoid speech distortion.

### C. Time-frequency comparison of the four algorithms

Similarly, a single test speech is selected here as an example. Figs. 3 and 4 are the time-domain comparison

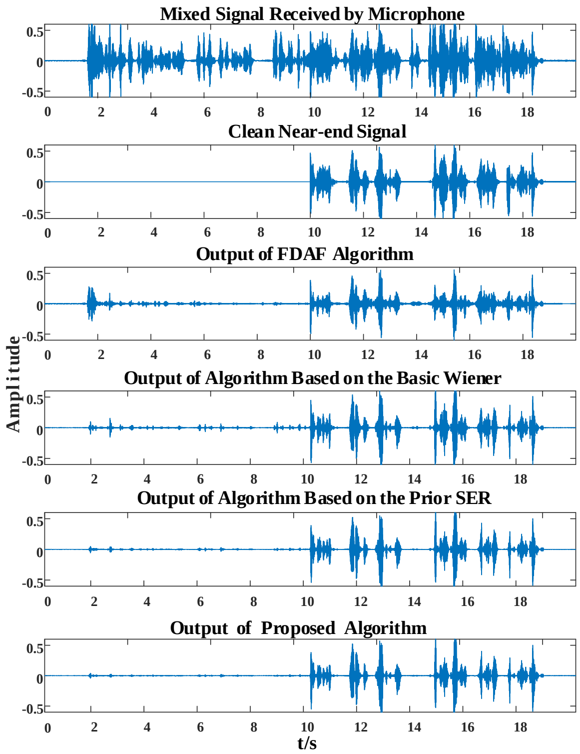


Fig. 3. Time-domain waveforms of the output of four echo suppression algorithms.

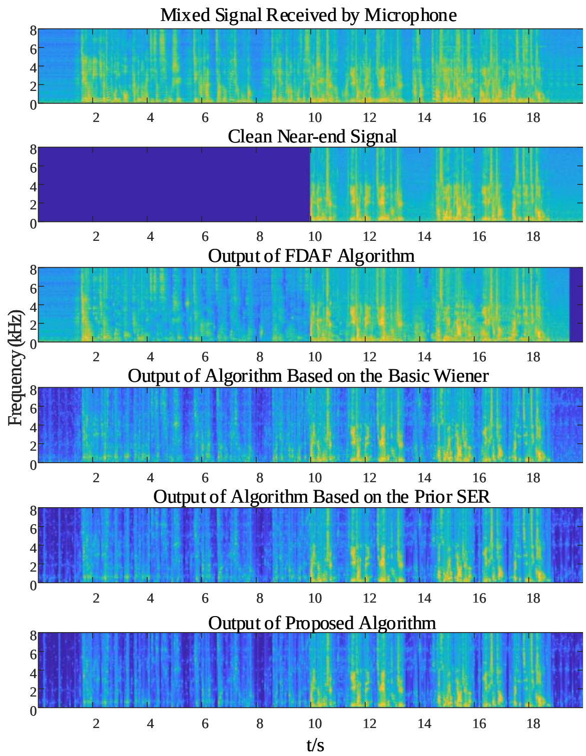


Fig. 4. Spectrograms of the output of four echo suppression algorithms.

diagram and speech spectrum comparison diagram of the test signal from four echo suppression algorithms, respectively.

From the time domain waveform of Fig. 3, we can see that the four algorithms were able to suppress echo. On the one hand, in comparison with traditional FDAF and the basic Wiener filtering, the echo suppression algorithm based on prior SER and the proposed AES algorithm can achieve significantly smaller residual echo at the single-far end. On the other hand, the convergence time of the three algorithms based on spectral correction was significantly less than that of the traditional FDAF in the first 3 s. The convergence time is defined as the time from the appearance of echo to the attenuation of echo being at -40 dB. The convergence time of FDAF, basic Wiener filtering, Wiener filtering based on SER, and the improved AES algorithm is 0.54, 0.2, 0.1, and 0.1 s, respectively.

It can be observed from Fig. 4 that the traditional FDAF algorithm still had obvious residual echo during the last 10 seconds that is in the double-talk state. The reason is that the system was easy to diverge due to the presence of near-end interference signals. Therefore, an double-talk detector was required to freeze the adaptive filter coefficients. The three algorithms based on spectral correction had no obvious residual echo, and do not need an additional control module.

In addition, the residual echo was effectively suppressed in the single-far end state, and the near-end speech was still retained when double-talk occurred in the improved algorithm.

### *D. ERLE indicators of the four algorithms*

In order to evaluate the superiority of the proposed method, the average ERLE indicators of four algorithms were compared in this section. We conducted 300 repeated experiments and the duration of the input speech is 20 s in each experiment. The signal in the first 10 s was only the far-end signal without the near-end voice, and the interaction state in the next 10 s was the double-talk. The experiment was repeated 300 times and Figs. 5 and 6 show the ERLE indicators of 300 far-end signals in the different communication states. The ERLE values of four algorithms are shown on the vertical y-axis, and the serial numbers of the test speeches are shown on the horizontal x-axis. It can be observed that the performance of the algorithm proposed in this paper is better than those of the other three algorithms. The ranking for the ERLE indicator was as follows: the proposed AES algorithm>the echo suppression algorithm based on SER>traditional Wiener filtering algorithm>FDAF method. The average ERLE indicators of 300 far-end signals of four algorithms are compared in Table 1. Compared with the echo suppression algorithm based on SER and the FDAF method, the improved algorithm achieves 4 dB and 16 dB ERLE improvement in the single-talk state, respectively. Even in the double-talk state in the last 10 s, the ERLE in the improved algorithm was higher than those in the other three algorithms. Furthermore, we can see that the ERLE in the single-talk state is higher than that in the double-talk state. When the near-end speech exists, the suppression curve is set to moderate attenuation to protect the quality of the near-end speech.

TABLE I
THE COMPARISON OF THE AVERAGE ERLE OF FOUR ALGORITHMS

| Method | ERLE in double-talk | ERLE in single-talk |
|---|---|---|
| FDAF | 6.5213 dB | 20.2031 dB |
| Algorithm based on Basic Wiener | 11.2657 dB | 26.2424 dB |
| Algorithm based on Prior SER | 13.1959 dB | 32.7254 dB |
| Improved AES algorithm | 13.7076 dB | 36.8510 dB |

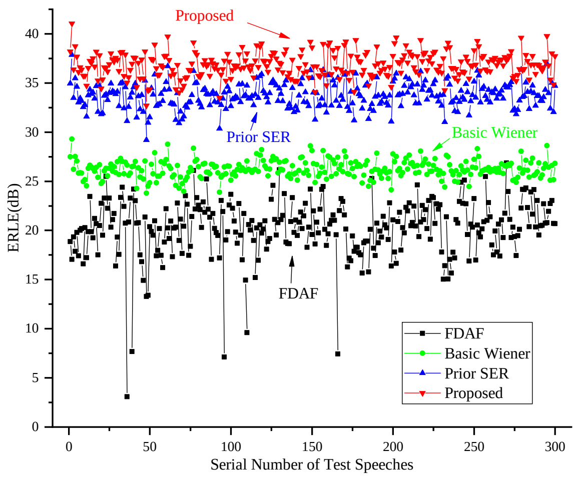


Fig. 5. The ERLE indicator of four algorithms in the single-talk state.

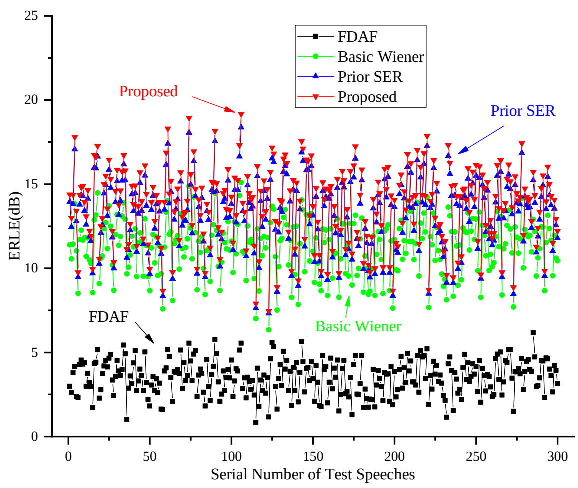


Fig. 6. The ERLE indicator of four algorithms in the double-talk state.

To evaluate the robustness of the system to the sudden change of the acoustic path, we assumed that the sudden change of the acoustic path occurred at 10 s. In the simulation experiment, the original path was multiplied by a coefficient $\tau$ from 10 s to simulate the sudden change in the acoustic path, $\tau = -2$ in this paper. We selected a test speech to conduct the experiment, and the duration of the test signal is 20s (320000 samples). Fig. 7 shows the ERLE indicators of the four algorithms. As shown in Fig. 7, the residual echo of FDAF algorithm increased rapidly and the corresponding ERLE decreased significantly when the echo path changes suddenly. The echo appeared at 11.5s. Then, the re-convergence process was completed within about 1 second, and the ERLE increased. Fig. 7 indicates that only the ERLE of the FDAF method immediately decreased and the ERLE of the three spectral correction schemes did not significantly deteriorate when the echo path suddenly changed. Moreover, the improved algorithm in this paper has better ERLE than the other three algorithms, showing that the spectral correction scheme based on the acoustic coloration effect had high robustness in the acoustic path mutation.

## V. Conclusion

In this paper, we used the basic Wiener filtering echo cancellation algorithm as the algorithm framework. Compared with the scheme based on the adaptive filter, these algorithms have ultra-low computational complexity and fast initial convergence characteristics. Firstly, a Wiener filter echo suppression scheme based on prior SER estimation was studied and it was extended to the parametric Wiener filter scheme. Then, we proposed a parameter adaptation strategy based on posterior SER and double-talk decision factors. The effectiveness of the proposed algorithm was verified through simulation experiments and the performance was compared with the FDAF, the basic Wiener filter algorithm, and the scheme based on prior SER. The simulation results show that the introduction of a parameter adaptation strategy based on posterior SER and double-talk decision factors in the parametric Wiener filter can effectively suppress the echo. Compared with the FDAF algorithm, the proposed algorithm improved ERLE by 16 dB with only 1/10th computational complexity. It also reduced the initial convergence time from 0.5 s to 0.1 s, and has high robustness to path mutation. Furthermore, the ERLE in the proposed algorithm was increased by 4 dB in comparison with that in the traditional scheme based on prior SER. Future studies would be focused on the multi-channel and low-complexity echo cancellation algorithms.

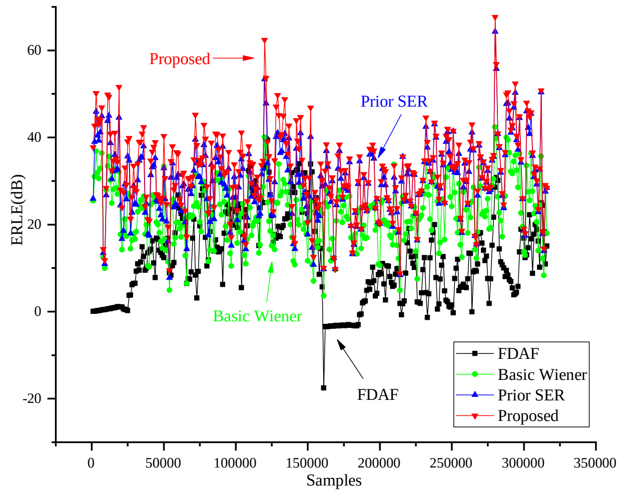


Fig. 7. The ERLE indicators of four algorithms in the path mutation at the middle time period.

## References

[1] Y. Zhou, C. Zheng, and X. Li, "A new robust gradient lattice ladder adaptive filtering algorithm for stereophonic acoustic echo cancellation," *Acta Acustica*, vol. 35, no. 2, pp. 223–229, 2010.

[2] L. T. T. Tran, H. Schepker, S. Doclo, H. H. Dam, and S. Nordholm, "Proportionate nlms for adaptive feedback control in hearing aids," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 211–215.

[3] J. Kapoor, G. Mishra, A. Pathak, and M. Rai, "Analysis of bsc and awgn channel distortion effect on sound signal in active noise cancellation application," *Engineering Letters*, vol. 29, no. 3, pp. 926–930, 2021.

[4] W. Chen, H. Sun, J. Zhu, G. Zhu, and M. Zhang, "A method for detecting target echo in reverberation based on fractional fourier transform," *Acta Acustica*, vol. 29, no. 4, pp. 401–412, 2010.

[5] K. Lee, Y. Baek, and Y. Park, "Nonlinear acoustic echo cancellation using a nonlinear postprocessor with a linearly constrained affine projection algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 9, pp. 881–885, 2015.

[6] B. Liao and Q. Xiang, "Discrete-time noise-suppressing zhang neural network for dynamic quadratic programming with application to manipulators," *Engineering Letters*, vol. 25, no. 4, pp. 431–437, 2017.

[7] M. Bekrani, A. W. H. Khong, and M. Lotfizad, “A linear neural network-based approach to stereophonic acoustic echo cancellation,” *IEEE Transactions on Audio Speech Language Processing*, vol. 19, no. 6, pp. 1743–1753, 2011.
[8] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, “Integrated echo and noise canceler for hands-free applications,” vol. 49, no. 3, pp. 188–195, 2002.
[9] H. Zhang and D. L. Wang, “Deep learning for acoustic echo cancellation in noisy and double-talk scenarios,” in *Proc. Interspeech 2018*, 2018, pp. 3239–3243.
[10] H. Huang, C. Hofmann, W. Kellermann, J. Chen, and J. Benesty, “A multiframe parametric wiener filter for acoustic echo suppression,” in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–5.
[11] Y. Tong and Y. Gu, “Acoustic echo suppression based on speech presence probability,” in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 35–38.
[12] P. Meyer, S. Elshamy, and T. Fingscheidt, “A multichannel kalman-based wiener filter approach for speaker interference reduction in meetings,” in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 451–455.
[13] J. Lin, S. Niu, Z. Wei, X. Lan, and K. Wang, “Speech enhancement using forked generative adversarial networks with spectral subtraction,” in *Proc. Interspeech 2019*, 2019, pp. 3163–3167.
[14] E. Verteletskaya and B. Simak, “Noise reduction based on modified spectral subtraction method,” *IAENG International Journal of Computer Science*, vol. 38, no. 1, pp. 82–88, 2011.
[15] S. Lefkimmiatis and P. Maragos, “Optimum post-filter estimation for noise reduction in multichannel speech processing,” in *14th European Signal Processing Conference*, 2006, pp. 1–5.
[16] R. S. Pujar, “Wiener filter based noise reduction algorithm with perceptual post filtering for hearing aids,” *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 7, pp. 69–81, 2019.
[17] C. Avendano, “Acoustic echo suppression in the stft domain,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 175–178.
[18] M. Toyoda, T. Furukawa, and D. Takahashi, “Study on echo suppression effect and coloration due to periodic-type diffusers,” *Applied Acoustics*, vol. 70, no. 5, pp. 722–729, 2009.
[19] F. Wallin and C. Faller, “Perceptual quality of hybrid echo canceler/suppressor,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, no. 1, 2004, pp. 157–160.
[20] Tong and Y. Gu, “A modified a priori ser for acoustic echo suppression using wiener filter,” in *IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016, pp. 1–4.
[21] V. Botchev, “Speech enhancement: theory and practice (2nd ed.),” *Computing reviews*, vol. 54, no. 10, pp. 604–605, 2013.
[22] Y. S. Park and J. H. Chang, “Frequency domain acoustic echo suppression based on soft decision,” *IEEE Signal Processing Letters*, vol. 16, no. 1, pp. 53–56, 2008.
[23] C. Faller and C. Tournery, “Estimating the delay and coloration effect of the acoustic echo path for low complexity echo suppression,” in *Proc. Intl. Works. on Acoust. Echo and Noise Control (IWAENC)*, 2005, pp. 1–4.
[24] J. Lei, “Consider of issues and countermeasures about computer-aided mandarin proficiency test,” *Journal of Wuhan University of Technology*, vol. 32, no. 13, pp. 160–163, 2010.