Received March 15, 2019, accepted April 2, 2019, date of publication April 12, 2019, date of current version May 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910858

An Improved Practical State-Space FDAF With Fast Recovery of Abrupt Echo-Path Changes

TAO JIANG[1], RUIYU LIANG[1,2], (Member, IEEE), QINGYUN WANG[2], CAIRONG ZOU[1], AND CONG LI[3]

[1]School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[2]School of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China
[3]School of Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

Corresponding author: Ruiyu Liang (lly1711@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871213 and Grant 61673108, and in part by the Six Talent Peaks Project in Jiangsu Province under Grant 2016-DZXX-023.

ABSTRACT A variable step-size echo cancellation algorithm in the frequency domain based on soft-decision and status information of the filter is established for enhancing tracking capability. The first-order Markov model is initially used to represent the time-variant acoustic path. In the meantime, the step size of the filter is replaced by the Kalman gain. Afterward, the magnitude squared coherence function of the reference signal and error signal are calculated as the filter status. The status of the filter is then mapped to a soft-decision value and is used to weigh the channel transmission parameter. Finally, the acoustic path is tracked by the weighted channel transmission parameter, which facilitates the reconvergence of the adaptive filter. The theoretical and experimental analysis is demonstrated on the validity of the algorithm. Compared to similar algorithms, the proposed algorithm has less misalignment but can also converge faster during abrupt echo-path changes. Compared to the state-space frequency domain adaptive filters (SS-FDAF) algorithm, the convergence time is 44% shorter. In addition, the amount of misalignment decreases by 8 dB. Moreover, the filter update time only increases by 15.38%.

INDEX TERMS Echo cancellation, state-space FDAF, echo-path changes, tracking capability.

I. INTRODUCTION

Acoustic echo is an inevitable problem for duplex intelligence speech interaction terminals. Using an adaptive filter is considered to be a regular method to estimate the acoustic path of loudspeaker-enclosure-microphones (LEM). Such algorithms in the time domain include normalized least mean square (NLMS [1]), affinity propagation (AP [2]) and recursive least squares (RLS [3]). However, the convergence speed of these algorithms is usually inefficient in the case of long echo paths. As such, NLMS in the frequency domain is more conventional among those algorithms when designed for limited resource speech terminal devices. The step-size is a critical parameter for the echo cancellation algorithm to control the adaptation of the filter. A large number of variable step-size solutions [4], [5] is proposed to solve the contradiction [6], [7] between the convergence rate and the misalignment of the fixed step-size frequency domain adaptive filters (FDAF) algorithm. These solutions mainly concentrate on how the filter can adaptively update its step size according to its convergence. The larger updated step is used to achieve fast convergence during the transient state of the filter, whereas a smaller step is used to achieve low misalignment.

There have been two types of variable step size strategies proposed in the last decade. One is based on the error signal [8]–[13]. In [8], [13], the step size is mapped to weighted factors by a nonlinear function. Hence, it can vary with error signal. Obvious improvements can be seen in both convergence rate and misalignment via this method. However, improvement in the recovery performance of the system is negligible. Moreover, this type of algorithm is modified by the combination of a priori error and a posteriori error. In [9], the step size changes with the self-correlation coefficient ratio of the previous two types of error. At the initial convergence stage, the convergence rate increases. Nevertheless, for the input of a nonstationary signal, the convergence depth is not ideal. To achieve further progress in the control of

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Angiulli.

misalignment, a shrinkage denoising method is applied on the evaluation of a priori error, which is based on a risk function that combines a priori error with a posteriori error [10]–[12]. Overall, the tracking rate during reconvergence and robustness during double talk are not as good as expected, although this type of method can further decrease misalignment, while efficiently maintaining the convergence rate.

The other method is based on the state-space of the acoustic path. It mainly focuses on approximating a variable acoustic path via a first-order Markov model. Meanwhile, it applies classic filtering theory to update the adaptive filter coefficient [14]. In [15], [16], the frequency domain Kalman filter is separated and fitted into the FDAF model, of which the step size is replaced by Kalman gain. This method achieves an excellent balance among convergence rate, misalignment and robustness during double talk. This is a DTD (double talk detection) free [17] robust algorithm that aims to minimize misalignment. The filter keeps updating at an approximately optimal step size. However, it can freeze when double talk occurs. Additionally, [18], [19] proposed a low-energy consumption method. It does not have a redundant parameter configuration, but the confliction between tracking ability of path variance and misalignment is non-negligible. The state-space model has a property in which the system's misalignment decreases as a greater transmission parameter is selected. However, the tracking ability of the filter deteriorates when an abrupt path change occurs; a deadlock could even occur in the filter. Only by setting the channel transmission parameter to a relatively small value can the system efficiently track the abrupt echo-path, but in sacrifice of misalignment. To solve this problem, a fixed step size shadow filter was presented in [5], [20]–[23]. This method can reset the filter via an echo-path change detector. The strategy can achieve excellent balance between tracking ability and misalignment but requires an increase in calculation and storage resources.

An improved state-space FDAF algorithm (BFS-SS-FDAF) is proposed based on the filter status information and soft decision. Under the circumstance of adding limited calculations, the algorithm can be maintained at low misalignment and guarantee fast convergence. It also has improved tracking ability during abrupt changes in acoustic echo. First, the structure of the state-space FDAF algorithm is established. Meanwhile, Kalman gain is used for updating the step size of the filter, which facilitates the filter to reach the optimal step size through iteration. Next, a filter status decision factor is proposed to detect abrupt changes in the echo-path. Meanwhile, nonlinear soft-decision functions and a correlation function are used to integrate the status information into a path modeling parameter. Then, the steady state of reconvergence is analyzed again. Additionally, the defects of the SS-FDAF and the effectiveness of the proposed algorithms are theoretically explained. Then, experiments are designed to compare and verify the difference in performance between the proposed and the original algorithms. Compared to references [5], [20], [21], the status information of the filter is integrated into the process of path modeling via soft decision, rather than adding another shadow filter. Hence, processing time consumption is only increased by 15.38%.

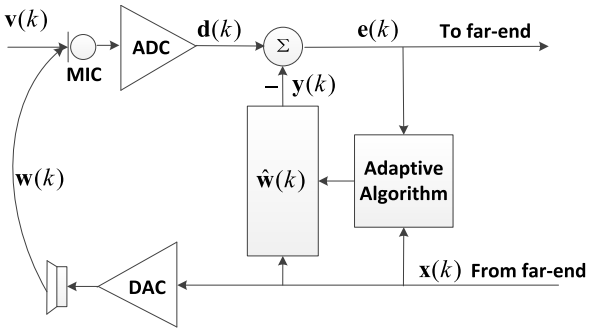


FIGURE 1. Basic echo canceller model.

II. STATE-SPACE FDAF

Fig. 1 shows the single-microphone and single-speaker echo canceller model. The microphone input signal $d(n)$ is represented as:

$$d(n) = x^T(n)w(n) + v(n) \tag{1}$$

Here, $x(n)$ is the reference signal; $x(n) = [x(n), \ldots, x(n - L + 1)]^T$. L is the signal length; and $v(n)$ is the near end signal, which contains background noise and near end speech. $w(n)$ and $x(n)$ have finite lengths L due to limited calculation resources. $w(n) = [w_0(n), \ldots, w_{L-1}(n)]^T$ can be viewed as a first order Markov model [16], which is expressed as:

$$\mathbf{W}(k+1) = T_A\mathbf{W}(k) + \Delta\mathbf{W}(k) \tag{2}$$

Here, $\mathbf{W}(k) = [W_0(k), \ldots, W_{2L-1}(k)]^T = \mathbf{F}[\mathbf{w}^T, \mathbf{0}_{1\times L}]$ and $\mathbf{F}$ represents Fourier matrices. T_A represents the path transmission parameter and $\Delta\mathbf{W}(k)$ is the process noise with a mean value of 0.

The error signal can be written as:

$$\mathbf{E}(k) = \mathbf{D}(k) - \mathbf{G}X(k)\hat{\mathbf{W}}(k) \tag{3}$$

Here, $\mathbf{G}$ is a constraint factor that guarantees the consistencies of the time and frequency domains.

$$\mathbf{G} = \mathbf{F}\begin{bmatrix} \mathbf{I}_L & 0 \\ 0 & 0 \end{bmatrix}\mathbf{F}^{-1} \tag{4}$$

The universal filter updating equation is

$$\hat{\mathbf{W}}(k+1) = \hat{\mathbf{W}}(k) + \boldsymbol{\mu}(k)\mathbf{X}^H(k)\mathbf{E}(k) \tag{5}$$

where $\boldsymbol{\mu}(k)$ is the update step size, and the value in a typical FDAF model is:

$$\boldsymbol{\mu}(k) = \alpha\boldsymbol{\Psi}_{XX}^{-1}(k) \quad 0 < \alpha < 1 \tag{6}$$

where α is a fixed step size and $\mathbf{\Psi}_{XX}(k)$ represents the energy spectrum density of far end reference signal. It can be written as:

$$\mathbf{\Psi}_{XX}(k) = \lambda\mathbf{\Psi}_{XX}(k-1) + (1-\lambda)\mathbf{X}(k)\mathbf{X}^H(k) \tag{7}$$

where λ is a forgetting factor and $0 < \lambda < 1$.

Transferring Kalman gain into step size $\boldsymbol{\mu}(k)$ [15], [16]

$$\mathbf{K}(k) = \frac{\mathbf{P}(k-1)\mathbf{X}^H(k)}{\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) + 2\mathbf{\Psi}_{vv}} \tag{8}$$

$$\boldsymbol{\mu}(k) = \frac{T\mathbf{P}(k-1)}{\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) + 2\mathbf{\Psi}_{vv}} \tag{9}$$

where $\mathbf{P}(k)$ is the variance of error, which can be expressed as:

$$\mathbf{P}(k+1) = T^2[\mathbf{I} - \frac{1}{2}\mathbf{K}(k)\mathbf{X}(k)]\mathbf{P}(k) + \mathbf{\Psi}_{\Delta\Delta}(k) \tag{10}$$

where $\mathbf{P}(k)$ represents the variance of the state estimate error. Here, $\mathbf{\Psi}_{\Delta\Delta}(k)$ is estimated using the method described in [15]:

$$\mathbf{\Psi}_{\Delta\Delta}(k) = (1-T^2)E[|\mathbf{W}(k)|^2] \tag{11}$$

III. IMPROVED ALGORITHM BASED ON FILTER STATUS INFORMATION STATE-SPACE FDAF (BFS-SS-FDAF)

A. TRANSMISSION PARAMETER CALCULATION

The connection between filter status information and an abrupt change in the echo-path is the major focus of this article. The relationships among abrupt changes in the echo-path and the correlation function $\gamma_{xe}(k)$ of the error signal $\mathbf{E}(k)$ with the reference signal $\mathbf{X}(k)$ are theoretically analyzed. Then, the correlation function $\gamma_{xe}(k)$ is integrated into the acoustic path modeling. This strategy solves the filter dead lock issue in SS-FDAF that occurs when there is an abrupt change in the acoustic path.

First, the error signal contains the correct residual echo of liner part $\xi(k)$, nonlinear part $\varepsilon(k)$ and near end signal $v(k)$. Generally, it is caused by abrupt changes in the echo-path and the nonlinearity of LEM.

$$e(k) = \xi(k) + \varepsilon(k) + v(k) \tag{12}$$

$$\xi(k) = y(k) - \hat{y}_r(k) \tag{13}$$

where $y(k)$ is echo in microphone signal and $\hat{y}_r(k)$ is the correctly estimated echo. Both nonlinear estimation error $\varepsilon(k)$ and near end signal $v(k)$ are uncorrelated with the reference signal $v(k)$ and linear residual echo $\xi(k)$. Hence:

$$|S_{xe}|^2 = S_{xx}S_{\xi\xi} \tag{14}$$

where S_{xy} represents the cross-power spectrum density of x and y. When $x = y$, it is seen as autopower spectral density. The calculation of frequency uses the exponential smoothing method:

$$\mathbf{S}_{x_k y_k} = \alpha\mathbf{S}_{x_k y_k} + (1-\alpha)\mathbf{X}(k)\mathbf{Y}^*(k) \tag{15}$$

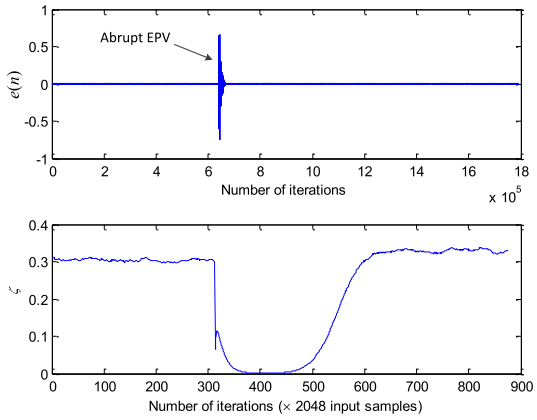


FIGURE 2. Error signal and decision factor when abrupt path change occurs.

where α is the smoothing factor. The correlation measurement factors of the reference and error signals are denoted by $\gamma_{xe}(k)$:

$$\gamma_{xe}(k) = \frac{S_{xe} \cdot S_{xe}^*}{S_{xx} \cdot S_{ee}} \tag{16}$$

Combining (16) with (14) yields

$$\gamma_{xe}(k) = \frac{S_{\xi\xi}}{S_{ee}} = 1 - \frac{1}{1 + \frac{S_{\xi\xi}}{S_{\varepsilon\varepsilon}+S_{vv}}} \tag{17}$$

As the echo path changes, $S_{\varepsilon\varepsilon}$ rapidly increases while $\gamma_{xe}(k)$ drops sharply. Afterward, if the filter can be converged normally, then $S_{\varepsilon\varepsilon}$ decreases gradually. Meanwhile, $S_{\xi\xi}$ increases gradually. Thus, $\gamma_{xe}(k)$ increases gradually. If the filter cannot converge appropriately, $S_{\varepsilon\varepsilon}$ continuously increases and $\gamma_{xe}(k)$ continuously decreases when deviating from Wiener solution. Hence, filter status information can be achieved from $\gamma_{xe}(k)$ during abrupt changes in the echo path. If the frequency spectrum distribution of speech and low frequency noise are considered at the same time, the mean of $\gamma_{xe}(k)$ in the specific frequency band is regarded as the decision factor:

$$\zeta = \frac{1}{I+1}\sum_{i=0}^{I}\gamma_{xe}(f_i) \tag{18}$$

where $f_i = f_L + i(f_H - f_L)/I$, $[f_L, f_H]$ is the decision band range. Fig. 2 illustrates the performance of the error signals $e(n)$ and ζ during abrupt path variance (EPV). The simulation result is consistent with (17).

The reconvergence performance is correlated to T in (11) [20]. Therefore, the Sigmoid and nonlinear cut off functions are used to transfer the filter status factor ζ into the soft-decision factor SD. The range is constrained within $[T_{\min}, \eta]$.

$$SD = \max\{2\eta \cdot (\frac{1}{1+e^{-\beta\zeta}} - 0.5), T_{\min}\} \tag{19}$$

After the completion of initial convergence (ICflag $= 1$), the soft decision value SD is used for amending T_i, where i is

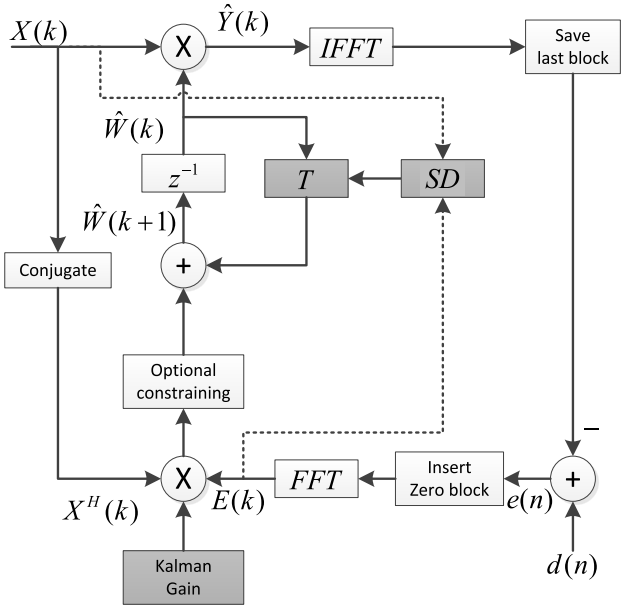


FIGURE 3. BFS-SS-FDAF algorithm diagram.

the frequency domain block number. In case of the effect of double talk, the soft decision value only weighs T_i when $\rho_{\hat{y}d}$ is greater than the threshold value.

$$T_i' = \begin{cases} T_i \cdot SD_i, & \text{ICflag} = 1 \& \rho_{\hat{y}d} > \tau \\ T_i, & \text{ICflag} = 0 \end{cases} \tag{20}$$

Here, the equation to evaluate $\rho_{\hat{y}d}$ is:

$$\rho_{\hat{y}d}(k) = \frac{S_{\hat{y}d} \cdot S_{\hat{y}d}^*}{S_{\hat{y}\hat{y}} \cdot S_{dd}} \tag{21}$$

By substituting (9) and (20) into (5), the BFS-SS-FDAF filter update equation is (22). A diagram is shown in Fig. 3.

$$\hat{\mathbf{W}}(k+1) = \begin{cases} \hat{\mathbf{W}}(k) + \dfrac{T \cdot SD \cdot \mathbf{P}(k-1)}{\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) + 2\boldsymbol{\Psi}_{vv}} \mathbf{X}^H(k)\mathbf{E}(k), \\ \quad \text{ICflag} = 1 \& \rho_{\hat{y}d} > \tau \\ \hat{\mathbf{W}}(k) + \dfrac{T \cdot \mathbf{P}(k-1)}{\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) + 2\boldsymbol{\Psi}_{vv}} \mathbf{X}^H(k)\mathbf{E}(k), \\ \quad \text{ICflag} = 0 \end{cases} \tag{22}$$

B. RECONVERGENCE ABILITY ANALYSIS

By substituting (11) and (20) into (10) and writing a steady-state solution:

$$P_i(\infty) = T_i'^2 (1 - \frac{1}{2} \frac{P_i(\infty)}{P_i(\infty) + \kappa_i}) P_i(\infty) + (1 - T_i'^2) E[|\mathbf{W}(k)|^2] \tag{23}$$

where $\kappa_i = E\{2|V_i(k)|^2\}/E\{2|X_i(k)|^2\}$. (23) can be rearranged into a second order equation, then solved as follows:

$$\widehat{P}_i(\infty) = \frac{1 - T_i'^2}{2 - T_i'^2} (1 - \frac{1}{\delta_i} + \sqrt{1 + \frac{1}{\delta_i^2} + \frac{2}{1 - T_i'^2} \frac{1}{\delta_i}}) \tag{24}$$

where $\delta_i = E[|\mathbf{W}(k)|^2]/\kappa_i$. Substituting (24) into (9) gives steady-state step size:

$$\mu(\infty) = \frac{\delta_i \widehat{P}_i(\infty)}{1 + \delta_i \widehat{P}_i(\infty)} = \frac{1}{1 + \frac{1}{\delta_i \widehat{P}_i(\infty)}} \tag{25}$$

To further understand the relationship between steady-state error $\widehat{P}_i(\infty)$ and T_i', the derivative of T_i' is taken:

$$\frac{\partial \widehat{P}_i(\infty)}{\partial T_i'} = \frac{-2T_i'}{(2 - T_i'^2)^2} \cdot \left(1 - \frac{1}{\delta_i} + \frac{\delta_i + \frac{1}{\delta_i} + \frac{T_i'^2}{1 - T_i'^2}}{\sqrt{1 + \delta_i^2 + \frac{2}{1 - T_i'^2} \delta_i^2}} \right) < 0 \tag{26}$$

From (26), misalignment $\widehat{P}_i(\infty)$ decreases when T_i' increases. Combining this with (24), as $T_i' \to 1$, $\widehat{P}_i(\infty) \to 0$. According to (25), $\mu(\infty) \to 0$. At this point, the filter is dead locked. When countering the abrupt path change, the performance of path tracking deteriorates. This is the reason why the filter is dead locked while abrupt path variance occurs in SS-FDAF.

However, in the proposed algorithm, the amended T_i' can be adaptively adjustable. As the path suddenly changes, the soft decision factor can effectively amend T_i'. Taking the derivative of ζ_i as follows,

$$\frac{\partial T_i'}{\partial \zeta_i} = \frac{2\eta\beta T_i}{(1 + e^{-\beta\zeta_i})^2} > 0 \tag{27}$$

From (27), ζ_i decreases before it increases. It follows the path transmission parameter T_i', which initially falls and then rises. Combining (26) with (27), the relationship between the steady-state solution $\widehat{P}_i(\infty)$ and filter status factor ζ_i can be analyzed:

$$\frac{\partial \widehat{P}_i(\infty)}{\partial \zeta_i} = \frac{\partial \widehat{P}_i(\infty)}{\partial T_i'} \cdot \frac{\partial T_i'}{\partial \zeta_i} < 0 \tag{28}$$

As described in (28), during the abrupt change in path, $\widehat{P}_i(\infty)$ increases as ζ_i decreases. Meanwhile, the step size increases rapidly as $\widehat{P}_i(\infty)$ increases according to (25), and the filter can rapidly converge to avoid dead lock. Then, its step size diminishes until reaching the steady state. Finally, $\mu(\infty) \to 0$, $\widehat{P}_i(\infty) \to 0$, and the misalignment maintains a relatively low level. Hence, theoretical analysis proves that the proposed algorithm has excellent tracking ability during abrupt changes at low levels of misalignment. This algorithm efficiently solves the dead lock issue due to abrupt path changes in SS-FDAF.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

In this section, the performance of the proposed method is evaluated via computer simulations and hardware tests in real-world scenarios. The proposed algorithm (BFS-SS-FDAF) is compared with state-space-FDAF [20].

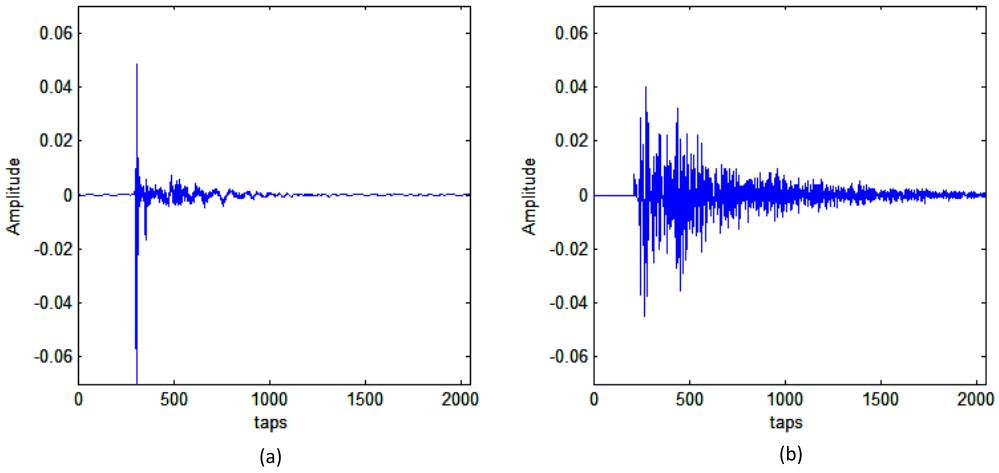


FIGURE 4. Impulse responses in the lab and conference room. (a) Laboratory acoustic path. (b) Conference room acoustic path.

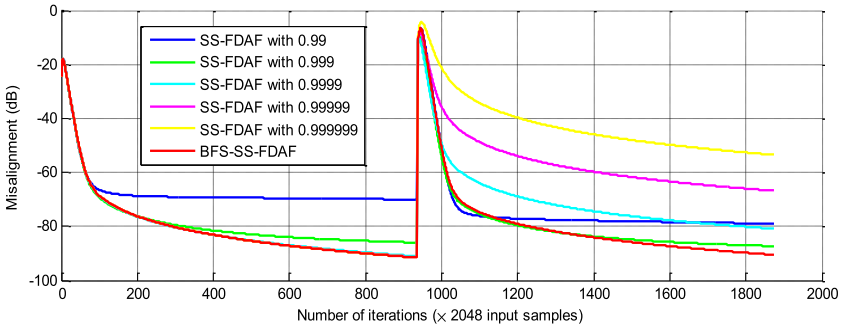


FIGURE 5. Misalignments of SS-FDAF with different T, the proposed algorithm. The single-talk scenario and the simulation path use the laboratory path.

The echo path is from the acoustic lab at Bar-llan University [24] and a series of data that were prerecorded in a conference room ($3m \times 4m \times 3m$). Fig. 4 shows two different room impulse responses, the impulse response of lab (a) and that of conference (b) room. The echo tail length is set to be 128 ms and the sample rate is 16 kHz. In addition, the input signal uses Gaussian white noise. To simulate the echo-path changes, at the middle of the experiment, the echo-path coefficients are multiplied by -4 to model an abrupt echo-path change. The values of parameter T in state-space FDAF [20] (SS-FDAF) are set to be 0.99, 0.999, 0.9999, 0.99999 and 0.999999. For the proposed algorithm, $\alpha = 0.93$, $\beta = 50$, $T_{\min} = 0.99$, $\eta = 1$ and $\tau = 0.93$. Moreover, the soft decision band is restricted within $f_L = 200Hz$ and $f_H = 3400Hz$. The main speech component of a teleconference system is mostly focused on this frequency band. Additionally, in DSP evaluation boards, the speaker's frequency response has great linearity within that frequency band.

The performance indicator uses filter misalignment (Mis). Less misalignment indicates that the estimated parameter of the adaptive filter is closer to the real room impulse response. It also means that a greater echo canceller performance is achieved. The misalignment is calculated as:

$$Mis = 10\log_{10}\frac{\|\mathbf{h}-\mathbf{w}\|^2}{\|\mathbf{h}\|^2} \tag{29}$$

In the real test environment, the actual room impulse and near end signal during double talk period cannot be captured. Hence, a 20 s far-end test is performed first, followed by another 20 s double talk test. In addition, echo cancellation performance at the far end in the first 20 s can be evaluated by two indices. One is MSE (mean square error). The lower the MSE, the better the algorithm. The equation is expressed as:

$$MSE = e^2(n) \tag{30}$$

The other indicator is echo return loss enhancement (ERLE), which is calculated by:

$$ERLE = 10\lg\frac{d^2(n)}{e^2(n)} \tag{31}$$

where $e(n)$ is error signal and $d(n)$ is expected signal. A greater value of ERLE means that greater echo canceller performance is achieved.

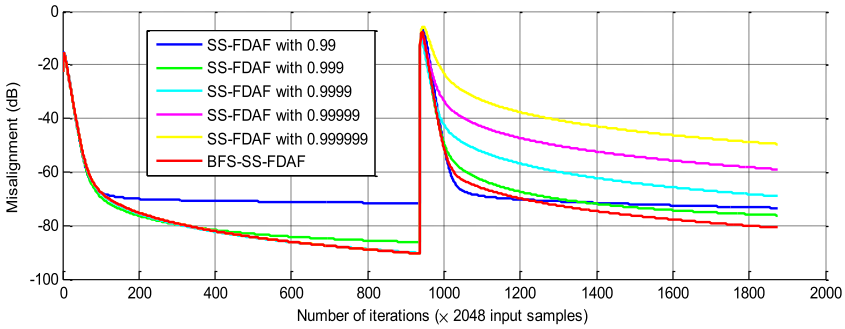


FIGURE 6. Misalignments of SS-FDAF with different T, the proposed algorithm. The single-talk scenario and the simulation path use the Conference room path.

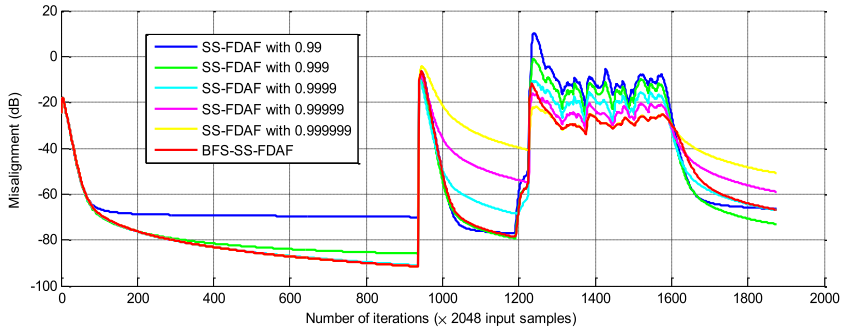


FIGURE 7. Misalignments of SS-FDAF with different TT, the proposed algorithm. The double-talk scenario and the simulation path use the Lab path.

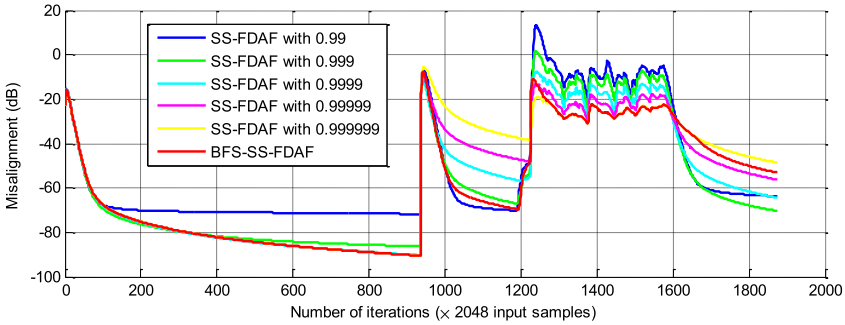


FIGURE 8. Misalignments of SS-FDAF with different TT, the proposed algorithm. The double-talk scenario and the simulation path use the Conference room path.

B. SINGLE FAR-END AND DOUBLE-TALK PERFORMANCE EVALUATION

Figs. 5 and 6 show the performance of the initial convergence at a single far end and path variance, in lab and conference rooms, respectively. The simulation results indicate that misalignment decreases as the value of T approaches 1. However, the tracking ability deteriorates during abrupt path changes. A comparison of the proposed algorithm BFS-SS-FDAF and SS-FDAF is shown below:

Misalignment of initial convergence:
$SS-FDAF(0.99) > SS-FDAF(0.999) > SS-FDAF(0.9999) \approx SS-FDAF(0.99999) \approx SS-FDAF(0.999999) \approx \mathbf{BFS-SS-FDAF}$

Convergence rate during abrupt path change:
$SS-FDAF(0.99) > \mathbf{BFS-SS-FDAF} > SS-FDAF(0.999999) > SS-FDAF(0.99999) > SS-FDAF(0.9999) > SS-FDAF(0.999)$

Misalignment of reconvergence:
$SS-FDAF(0.999999) > SS-FDAF(0.99999) > SS-FDAF(0.9999) \approx SS-FDAF(0.99) > SS-FDAF(0.999) > \mathbf{BFS-SS-FDAF}$

The simulation result is consistent with the theoretical results in Section 2.2. BFS-SS-FDAF efficiently tracks path variance during abrupt changes. In the meantime, it maintains low misalignment. Specifically, compared to SS-FDAF (0.9999), the tracking speed is shortened by 44.4% and misalignment decreases by 8 dB.

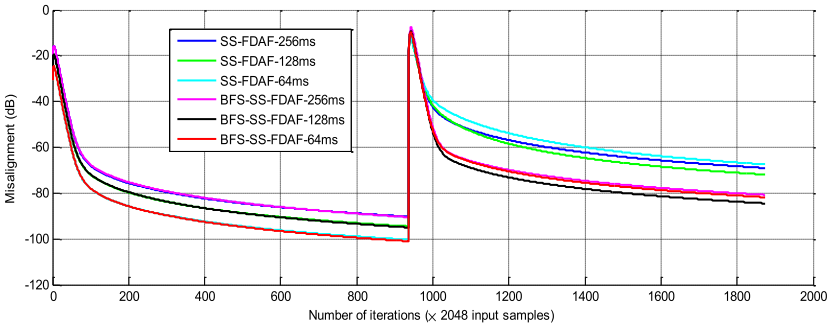


FIGURE 9. Misalignments of SS-FDAF and BFS-SS-FDAF with different echo tail lengths.

Figs. 7 and 8 represent misalignments with double talk, of SS-FDAF with different TT, of the proposed algorithm in the lab and conference rooms, respectively. Double talk is inserted after initial convergence. The simulation verifies that SS-FDAF is a DTD-free solution. During double talk, the proposed solution has a 5 dB improvement in misalignment. However, the convergence rate becomes slightly slower after double talk.

C. COMPARISON OF DIFFERENT ECHO TAIL LENGTHS

Echo paths with different tail lengths (64 ms, 128 ms, 256 ms) are obtained from the echo-path data sets. The test results are shown in Fig. 9. TT is 0.9999. According to Fig. 9:

Misalignment of initial convergence:
$SS-FDAF(256ms) \approx \mathbf{BFS-SS-FDAF}(256ms) > SS-FDAF(128ms) \approx \mathbf{BFS-SS-FDAF}(128ms) > SS-FDAF(64ms) \approx \mathbf{BFS-SS-FDAF}(64ms)$
Misalignment of reconvergence: $SS-FDAF(64ms) > SS-FDAF(256ms) > SS-FDAF(128ms) > \mathbf{BFS-SS-FDAF}(256ms) \approx \mathbf{BFS-SS-FDAF}(64ms) > \mathbf{BFS-SS-FDAF}(128ms)$

According to the experiment above, the proposed algorithm has a better convergence depth with three different echo tail lengths (64 ms, 128 ms, 256 ms) than SS-FDAF. In addition, when the sample rate is 16 k and block size is 2048, the algorithm is optimal for a 128 ms acoustic path. This is because when the sample rate is 16 k, a block size of 2048 matches the path length of 128 ms. Misalignment decreases as the echo tail length becomes longer than filter modeling length. Therefore, in a practical application, the filter block size should be suitably selected according to the real echo tail length.

D. PERFORMANCE TEST IN REAL SCENE

The proposed algorithm is applied on TI-DSP evaluation boards equipped with a single microphone and single speaker, as shown in Fig. 10. The connection of the algorithm and transparent transmission modules is demonstrated in Fig. 11 which also displays different types of gain in the system. The AEC input (AecIn) and output (AecOut) signals are synthesized to a two-channel stereo by USB audio, which can be uploaded to a PC. This setup facilitates the further evaluation of algorithm performance. The transmit channel

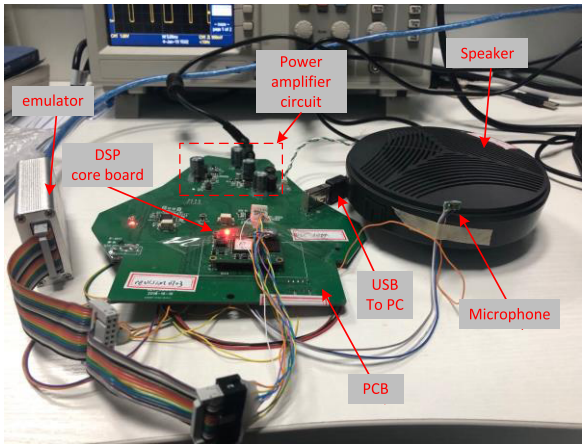


FIGURE 10. Audio conference hardware system.

gain is set to 0 dB. The reception channel gain is set to −3 dB in case the speaker enters the nonlinear region. The total testing time is 40 s. The first 20 s re for a single far end test during initial convergence, and the remaining 20 s are for double talk. The far end continues playing a female voice, and near the end, plays a male voice, in order to subjectively evaluate the quality of double talk.

Figs. 12 and 13 show two indices that are single far-end MSE and ERLE. According to these two indices, MSE is lower than in the original FDAF method, and ELER is also higher than in the original value, by an average of 3 dB. These two indices indicate that the proposed algorithm has a better convergence depth. The convergence depth is enhanced, especially in the first 4 s. Additionally, the residual echo is obviously weaker. For further comparison, the micro input signal (MicIn) and time domain waveform and spectrum are plotted.

Fig. 14 is a complete time domain waveform and Fig. 15 shows the associated spectrogram. Comparing the time domain waveform and spectrum in the first 4 s (the initial stage of convergence), residual echo is clearly less than that associated with the original algorithm. Furthermore, during the double talk stage in the next 20 s, the remaining audible echo of the proposed algorithm is obviously less than that of FDAF. This indicates that the filter can be frozen at the

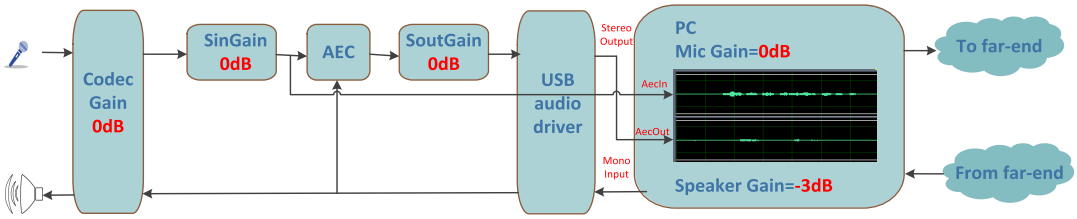


FIGURE 11. Algorithm module and transparent transmission module connection diagram with gain setup.

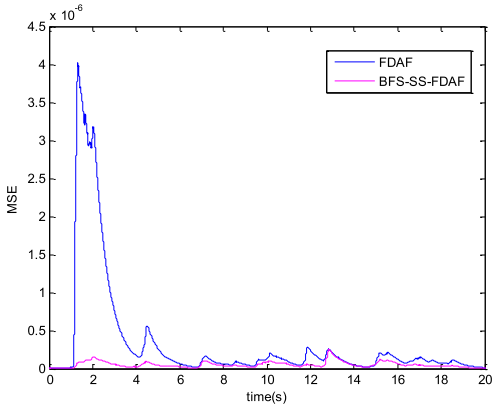


FIGURE 12. Single far end MSE performance in the first 20 s.

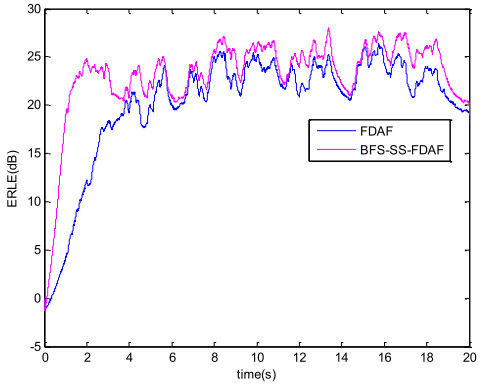


FIGURE 13. Single far end ERLE performance in the first 20 s.

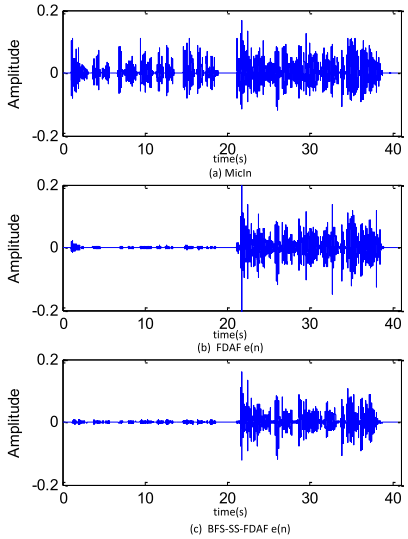


FIGURE 14. FDAF and BF-SS-FDAF output waveforms.

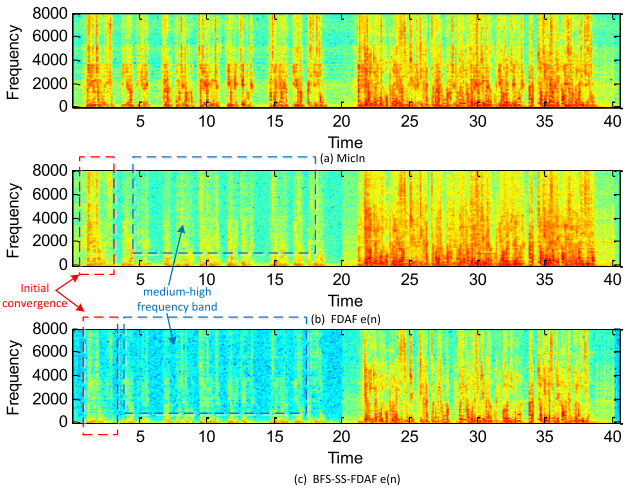


FIGURE 15. FDAF and BFS-SSFDAF output spectra.

beginning of double talk. For the original algorithm, echo continuously exists because the filter is not frozen in time.

E. TIME CONSUMPTION EVALUATION

The real-time performance and low power dissipation are key factors of voice terminal devices such as conference phones, intelligent sound and hearing aids. Hence, in this section, the number of real multiplication operations of the proposed algorithm (BFS-SS-FDAF), FDAF, SS-FDAF and that of Yang [20] with a shadow filter (state-space double frequency domain adaptive filters, SS-DFDAF) are compared in Tab. 1. The number of real multiplication operations of the four algorithms are expressed by FFT length and filter length. For FDAF, FFT length usually doubles the filter length ($N_1 = 2M$). For SS-DFDAF, the master filter length is M, and the slave filter length is $M/2$. Hence, the number of FFT points becomes $N_2 = M$. Moreover, Fig. 16 shows the average time used to update once the filter is applied for each algorithm. Block lengths are 512, 1024 and 2048, respectively. The SS-FDADF master filter uses a block length of 2048, whereas the slave filter uses a 1024 block size. The simulation is conducted using MATLAB2014b, with

TABLE 1. Real multiplication of different methods.

METHOD	FFT LENGTH	FILTER LENGTH
FIXFDAF	$5N_1 \log_2 N_1 + 8N_1$	$10M \log_2 M + 26M$
SS-FDAF	$5N_1 \log_2 N_1 + 8N_1 + 12N_1$	$10M \log_2 M + 50M$
SS-DFDAF	$5N_1 \log_2 N_1 + 8N_1 + 12N_1 + 5N_2 \log_2 N_2 + 8N_2$	$15M \log_2 M + 58M$
BFS-SS-FDAF	$5N_1 \log_2 N_1 + 8N_1 + 12N_1 + 12N_1$	$10M \log_2 M + 74M$

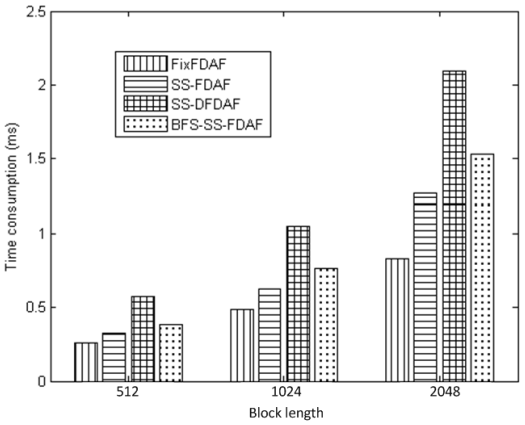


FIGURE 16. Time consumption of different methods.

an Intel Core i5-4900 processor and 16G RAM. As shown in Fig.'16, the longer the filter length, the more filter updating time is taken. When the filter length is 2048, Fix-FDAF needs 0.8 ms to update each time. 1.3 ms is required for SS-FDAF and 2.1 ms is required for SS-DFDAF. However, the proposed BFS-SS-FDAF only takes 1.5 ms to update. The test result indicates that SS-DFDAF requires 61.53% more time, but the time consumption of the proposed method only increases by 15.38%. Therefore, this algorithm is more suitable for scenarios that require more from real-time performance.

V. CONCLUSION

To solve the contradiction between the misalignment and tracking ability of the state-space FDAF, an improved state-space FDAF method is proposed. This method is based on filter status and soft decision technology. The proposed algorithm improves both the tracking ability following abrupt path changes and the steady-state misalignment performance. Additionally, theoretical analysis explains the filter deadlock issue that occurs during abrupt path changes and demonstrates the validity of the proposed method. The simulation experiment verifies that the proposed algorithm has excellent tracking ability during abrupt changes at low levels of misalignment. The experimental results are consistent with the theoretical analysis. Most notably, the time complexity of the proposed algorithm has a negligible increase compared to the original algorithm. Therefore, a reasonable balance between misalignment, the tracking ability of convergence and time complexity is achieved in this paper. Research regarding the stereo state-space FDAF will be conducted in the future.

REFERENCES

[1] U. Mahbub and S. A. Fattah, "A single-channel acoustic echo cancellation scheme using gradient-based adaptive filtering," *Circuits, Syst., Signal Process.*, vol. 33, pp. 1541–1572, May 2014.
[2] Y.-S. Kim, J.-H. Song, S.-K. Kim, and S. Lee, "Variable step-size affine projection algorithm based on global speech absence probability for adaptive feedback cancellation," *J. Central South Univ.*, vol. 21, pp. 646–650, Feb. 2014.
[3] C. Elisei-Iliescu, C. Stanciu, C. Paleologu, J. Benesty, C. Anghel, and S. Ciochină, "Robust variable-regularized RLS algorithms," in *Proc. Hands-Free Speech Commun. Microphone Arrays*, 2017, pp. 171–175.
[4] Y. Aytar, C. Vondrick, and A. Torralba. (2016). "SoundNet: Learning sound representations from unlabeled video." [Online]. Available: https://arxiv.org/abs/1610.09001
[5] F. Yang, G. Enzner, and J. Yang, "Statistical convergence analysis for optimal control of DFT-domain adaptive echo canceler," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 5, pp. 1095–1106, May 2017.
[6] K. Shi and X. Ma, "A frequency domain step-size control method for LMS algorithms," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 125–128, Feb. 2009.
[7] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.
[8] M. Hamidia and A. Amrouche, "Improved variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation," *Digit. Signal Process.*, vol. 49, pp. 44–55, Feb. 2016.
[9] J.-C. Liu, X. Yu, and L. Hong-Ru, "A nonparametric variable step-size NLMS algorithm for transversal filters," *Appl. Math. Comput.*, vol. 217, pp. 7365–7371, May 2011.
[10] Z. A. Bhotto and A. Antoniou, "A family of shrinkage adaptive-filtering algorithms," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1689–1697, Apr. 2013.
[11] P. Wen and J. Zhang, "A novel variable step-size normalized subband adaptive filter based on mixed error cost function," *Signal Process.*, vol. 138, pp. 48–52, Sep. 2017.
[12] X. Wei, L. Zhu, J. L. Zhu, J. Hu, and H. Li, "A shrinkage variable step size for normalized subband adaptive filters," *Signal Process.*, vol. 129, pp. 56–61, Dec. 2016.
[13] Y.-G. Zhu, Y.-G. Li, S.-Y. Guan, and Q.-S. Chen, "A novel variable step-size NLMS algorithm and its analysis," *Procedia Eng.*, vol. 29, pp. 1181–1185, Jan. 2012.
[14] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 8, pp. 1539–1549, Aug. 2013.
[15] F. Kuech, E. Mabande, and G. Enzner, "State-space architecture of the partitioned-block-based acoustic echo controller," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 1295–1299.
[16] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Process.*, vol. 86, pp. 1140–1156, Jun. 2006.
[17] T. Imbiriba, J. C. M. Bermudez, J. Y. Tourneret, and N. J. Bershad, "A new decision-theory-based framework for echo canceler control," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4304–4314, Aug. 2018.
[18] C. Wu, X. Wang, Y. Guo, Q. Fu, and Y. Yan, "Robust uncertainty control of the simplified Kalman filter for acoustic echo cancelation," *Circuits Syst. Signal Process.*, vol. 35, pp. 4584–4595, Dec. 2016.

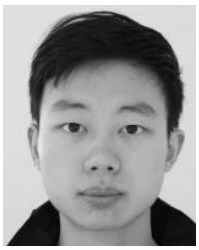
[19] L. M. Valero, E. Mabande, and E. A. P. Habets, "A state-space partitioned-block adaptive filter for echo cancellation using inter-band correlations in the Kalman gain computation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 599–603.
[20] F. Yang, G. Enzner, and J. Yang, "Frequency-domain adaptive Kalman filter with fast recovery of abrupt echo-path changes," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1778–1782, Dec. 2017.
[21] F. Yang and J. Yang, "Optimal step-size control of the partitioned block frequency-domain adaptive filter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 6, pp. 814–818, Jun. 2018.
[22] J. Guo, F. Yang, G. Enrner, and J. Yang, "Tracking analysis and improvement of broadband Kalman filter using the two-echo-path model as a rapid tunnel," in *Proc. 16th Int. Workshop Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 341–345.
[23] F. Yang, M. Wu, P. Ji, and J. Yang, "A robust variable step-size frequency-domain adaptive filter algorithm for acous-tic echo cancellation," in *Proc. 21st Int. Congr. Sound Vib.*, 2014, pp. 1–8.
[24] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. Int. Workshop Acoustic Signal Enhancement*, 2014, pp. 313–317.

TAO JIANG received the B.S. degree in electronic information engineering from the Nanjing Institute of Technology, Nanjing, China, in 2017. He is currently pursuing the M.S. degree in the field of signal processing with the School of Information Science and Engineering, Southeast University, Nanjing. He is interested in the fields of real-time speech signal processing and speech enhancement.

RUIYU LIANG (M'18) received the Ph.D. degree from Southeast University, China, in 2012. He is currently an Associate Professor with the Nanjing Institute of Technology, Nanjing, China. His research interests include speech signal processing and signal processing for hearing aids.

QINGYUN WANG received the M.S. degree in computer engineering, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2001 and 2011, respectively. She is currently a Professor with the Nanjing Institute of Technology, Nanjing. Her current research interests include acoustic signal processing, speech enhancement, and microphone array signal processing.

CAIRONG ZOU received the B.S., M.S., and Ph.D. degrees from Southeast University, China, in 1984, 1987, and 1991, respectively, all in electrical engineering. In 1992, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Concordia University, Canada. He is currently a Professor with Southeast University.



CONG LI received the B.Sc. degree in electrical systems from The University of Melbourne, Victoria, Australia, in 2018, where he is currently pursuing the master's degree in electronic and electrical engineering. He is interested in the fields of high-speed electronics, such as high-speed analogue circuits and bipolar junction transistors.