INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A Partitioned-Block Frequency-Domain Adaptive Kalman Filter for Stereophonic Acoustic Echo Cancellation

Rui Zhu[1], *Feiran Yang*[2], *Yuepeng Li*[1], *Shidong Shang*[1]

[1]Tencent Ethereal Audio Lab, Tencent Corporation, Beijing, China
[2]Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

raymondrzhu@tencent.com, feiran@mail.ioa.ac.cn, felixypli@tencent.com, simeonshang@tencent.com

Abstract

The rapid development of online video conferencing systems has caused renewed attention to the multi-channel recording and playback systems. Stereophonic acoustic echo cancellation (SAEC) is the key issue of this systems. This paper proposes an optimally designed partitioned-block frequency-domain Kalman filter (PBFDKF) algorithm for SAEC. We establish the frequency-domain observation equation using the overlap-and-save method and we use the first-order Markov model to describe the state equation. The exact PBFDKF algorithm is derived under the umbrella of Kalman filter theory and two fast implementations are then presented to reduce the complexity. The proposed algorithm is equivalent to the dual-channel partitioned-block frequency-domain gradient-based algorithm with optimum step-size control, and hence it exhibits very good convergence performance and is found to be robust to near-end interference without a double-talk detector. Extensive experiments in different SAEC conditions confirm the effectiveness of the proposed algorithm.

Index Terms: Stereophonic acoustic echo cancellation, Frequency-domain adaptive filter, Step-size control, Kalman filter

1. Introduction

The recent outbreak of coronavirus pandemic made it inconvenient for people to communicate face-to-face. This resulted in the popularization of online video conference platforms such as Skype, Zoom, and VooV Meeting. With the widespread use of video conferencing, participants want the experience of online meetings closer to the face-to-face communication. Therefore, algorithmic framework supporting high-fidelity multi-channel recording and playback system has attracted more and more attention. In a typical application scenario as shown in Fig. 1, the stereo signals captured by two far-end microphones are played through the stereo loudspeakers (SPK 0 and SPK 1) at the near end, and then they are picked up by the two near-end microphones (MIC 0 and MIC 1) due to the acoustic coupling. Therefore, stereophonic acoustic echo cancellation (SAEC) should be adopted for such online conference systems to reduce echos.

Stereo system can indeed provide more spatial information than the single-channel case. However, the SAEC problem in such a system is very challenging because the stereo signals are not only auto-correlated but also highly cross-correlated. This leads to the well-known non-uniqueness problem [1, 2]. An effective method for overcoming the non-uniqueness problem is to preprocess the stereo signals before they are sent to the loudspeakers [3–5]. Those solution can reduce the cross-correlation in certain degree, but they degrade the speech quality inevitably and even introduce audible distortion. Thus, the nonlinear preprocessing approaches are less successful and should not be adopted for high-quality video conferencing system.

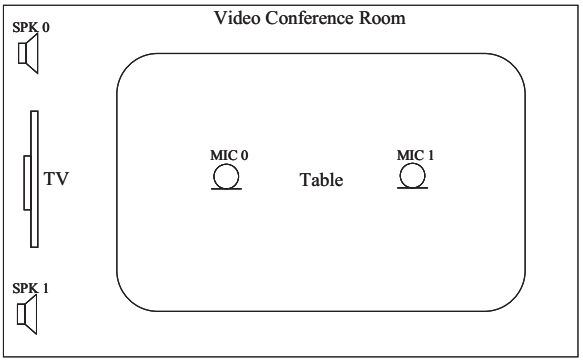

Figure 1: *Investigated meeting scenario with stereophonic playback and recording system.*

Another problem is the very slow convergence of the adaptive filters due to the ill-conditioned covariance matrix of the input signal. Many sophisticated adaptive filtering algorithms have been proposed to improve the convergence speed [6–8]. Considering that the room impulse response usually exceeds several thousand orders, the frequency domain adaptive filter (FDAF) algorithm has become a standard solution for acoustic echo cancellation (AEC) due to its lower complexity and better convergence performance [9]. However, it is well known that the step-size selection for the gradient-based algorithms is a very challenging problem. Therefore, the step-size control strategy that does not require an explicit double-talk detector is particularly welcome [10–12]. The Kalman filter algorithm has inherent decorrelation characteristics and its simplified form is equivalent to the gradient-based adaptive algorithm with the optimal step size. The Frequency-domain Kalman filter (FDKF) has been widely adopted in the field of audio signal processing since the pioneering work [13]. The multichannel state-space frequency-domain adaptive filter (MCSSFDAF) was proposed for multichannel AEC [14], which performs well in the presence of near-end interference and echo-path variability. A variationally-diagonalized version that ignores all the cross-channel terms of the MCSSFDAF was then presented to reduce the complexity [15]. Several approaches were also presented to further improve the convergece performance of the MCSSFDAF algorithm, e.g., in [16] and [17]. Although the DNN-based AEC solutions are favored, the design of the linear echo canceller is of high importance. For instance, the integra-

http://dx.doi.org/10.21437/Interspeech.2021-135

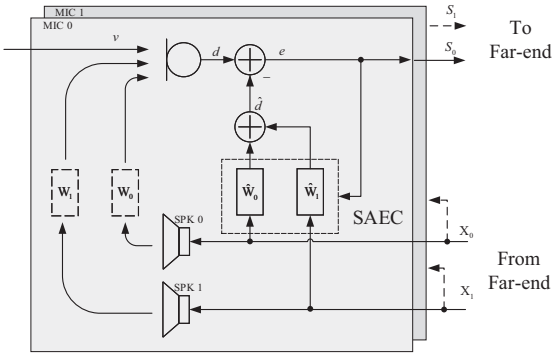

Figure 2: *Block diagram of SAEC.*

tion of a linear echo canceller and a DNN poster-filter is utilized in the top-ranking solution for single-channel AEC challenge [18, 19].

The standard frequency-domain adaptive algorithm introduces a two-frame delay [8, 11, 14], which is undesired for real-time video conferencing systems, especially when the room reverberation time is very long. To resolve this problem, we here propose a partitioned-block frequency-domain Kalman filter (PBFDKF) algorithm for SAEC. Using the overlap-and-save method, we establish the frequency-domain observation model. We then utilize a first-order Markov process to describe the echo path variation. The exact PBFDKF algorithm for SAEC is derived and two simplified forms are then presented to reduce the complexity. Finally, experiments in different SAEC scenarios are carried out to verify the performance of the proposed algorithm.

2. System model

2.1. Observation model

Fig. 2 illustrates the block diagram of stereo recording and playback system. The signal $d(n)$ captured by one of the near-end microphones at discrete time index n can be described by

$$d(n) = \sum_{i=0}^{1} \mathbf{x}_i^T(n)\mathbf{w}_i(n) + v(n), \tag{1}$$

where superscript $(\cdot)^T$ denotes transpose, $\mathbf{x}_i(n) = [x_i(n), \cdots, x_i(n-N+1)]^T$ is the input vector of channel i with length N, $\mathbf{w}_i(n) = [w_{i,0}(n), \cdots, w_{i,N-1}(n)]^T$ represents the unknown echo path between the i^{th} loudspeaker and one of the near-end microphones with length N, and $v(n)$ includes both the near-end speech and the background noise.

We now turn to the frequency domain. The weight vector $\mathbf{w}_i(n)$ is segmented into P smaller sub-blocks as $\mathbf{w}_{i,p}(k) = [w_{i,pL}(k), \cdots, w_{i,pL+L-1}(k)]^T$ with length $L = N/P$, where k denotes frame index. We transform $\mathbf{w}_{i,p}(k)$ into the frequency domain

$$\mathbf{W}_{i,p}(k) = \mathbf{F}\begin{bmatrix} \mathbf{w}_{i,p}(k) \\ \mathbf{0}_{L\times 1} \end{bmatrix}, \tag{2}$$

where $\mathbf{F}$ is the Fourier matrix of size $M \times M$ with $M = 2L$, and $\mathbf{0}_{L\times 1}$ represents an all-zero column vector with dimension $L \times 1$. The frequency-domain input matrix of p^{th} partition for the i^{th} channel is

$$\begin{aligned}\mathbf{X}_{i,p}(k) = \text{diag}\{\mathbf{F}[&x_i(kL-pL-L), \cdots, \\ &x_i(kL-pL+L-1)]^T\},\end{aligned} \tag{3}$$

where $\text{diag}\{\cdot\}$ creates a diagonal matrix from a vector. Using the overlap-and-save technique, we obtain the frequency-domain representation of (1)

$$\mathbf{D}(k) = \sum_{i=0}^{1}\sum_{p=0}^{P-1} \mathbf{G}_{01}\mathbf{X}_{i,p}(k)\mathbf{W}_{i,p}(k) + \mathbf{V}(k), \tag{4}$$

where $\mathbf{D}(k) = \mathbf{F}[\mathbf{0}_{1\times L}, d(kL), \cdots, d(kL+L-1)]^T$ and $\mathbf{V}(k) = \mathbf{F}[\mathbf{0}_{1\times L}, v(kL), \cdots, v(kL+L-1)]^T$ are the frequency-domain desired signal and noise signal vectors, respectively, and the windowing matrix $\mathbf{G}_{01} = \mathbf{F}\mathbf{Q}_{01}^T\mathbf{Q}_{01}\mathbf{F}^{-1}$ forces the last half of the corresponding time-domain vector of $\mathbf{X}_{i,p}(k)\mathbf{W}_{i,p}(k)$ to zero. We use the definition $\mathbf{Q}_{01} = [\ \mathbf{0}_L \quad \mathbf{I}_L\ ]$, where $\mathbf{0}_L$ and $\mathbf{I}_L$ denote the $L \times L$ zero and identity matrices, respectively. We can further rewrite (1) in a more compact matrix-vector multiplication form

$$\mathbf{D}(k) = \mathbf{X}(k)\mathbf{W}(k) + \mathbf{V}(k), \tag{5}$$

where $\mathbf{X}(k) = \mathbf{G}_{01}[\mathbf{X}_{0,0}(k), \cdots, \mathbf{X}_{0,P-1}(k), \mathbf{X}_{1,0}(k), \cdots, \mathbf{X}_{1,P-1}(k)]$ is the augmented input matrix, and $\mathbf{W}(k) = [\mathbf{W}_{0,0}^T(k), \cdots, \mathbf{W}_{0,P-1}^T(k), \mathbf{W}_{1,0}^T(k), \cdots, \mathbf{W}_{1,P-1}^T(k)]^T$ is the augmented weight vector composed of all the sub-block filters. Now, we have established the frequency-domain observation equation under the SAEC framework.

2.2. State-space model

We now discuss the establishment of the state-space model. In the real acoustic environment, the variability of the echo path is very complicated, and it is almost impossible to accurately describe this change. In [13], a first-order Markov model was adopted to describe the variability of the echo path, which has proven to be very effective

$$\mathbf{W}(k) = A\mathbf{W}(k-1) + \Delta\mathbf{W}(k), \tag{6}$$

where A is the transition parameter and $\Delta\mathbf{W}(k)$=$[\Delta\mathbf{W}_{0,0}^T(k), \cdots, \Delta\mathbf{W}_{0,P-1}^T(k), \Delta\mathbf{W}_{1,0}^T(k), \cdots, \Delta\mathbf{W}_{1,P-1}^T(k)]^T$ is the process noise vector of dimension $2PL \times 1$, which is a zero-mean random signal independent of $\mathbf{W}(k)$. The covariance matrix $\boldsymbol{\psi}_\Delta(k) = E[\Delta\mathbf{W}(k)\Delta\mathbf{W}^H(k)]$ includes $4P^2$ sub-matrices $\boldsymbol{\psi}_{\Delta,i,j,p,q}(k), 0 \le i,j \le 1, 0 \le p,q \le P-1$ of dimension $M \times M$, where $(\cdot)^H$ denotes Hermitian transpose. We further assume that the process noise between different channels is independent of each other, and thus $\Delta\mathbf{W}(k)$ can be approximated as a diagonal matrix:

$$\boldsymbol{\psi}_\Delta(k) \approx (1-A^2)\text{diag}\left\{\mathbf{W}(k) \odot \mathbf{W}^H(k)\right\}, \tag{7}$$

where $\odot$ represents the point-wise multiplication. Essentially, the transition parameter A and the energy of the echo path are jointly employed to describe the change of the echo path over time. If the noise covariance matrix can be accurately estimated, the model in (7) can well cope with the typical echo path changes even for a larger A [10].

3. Proposed algorithm

3.1. Accurate two-channel Kalman filter

Based on the observation equation (5) and state equation (6), we derive the exact two-channel PBFDKF algorithm as follows [20]

$$\mathbf{E}(k) = \mathbf{D}(k) - \mathbf{X}(k)\hat{\mathbf{W}}(k-1), \tag{8}$$

$$\mathbf{P}(k-1) = A^2[\mathbf{I}-\mathbf{K}(k-1)\mathbf{X}(k-1)]\mathbf{P}(k-2)+\boldsymbol{\psi}_{\Delta}(k), \tag{9}$$

$$\mathbf{K}(k) = \mathbf{P}(k-1)\mathbf{X}^H(k)[\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) + \boldsymbol{\psi}_v(k)]^{-1}, \tag{10}$$

$$\hat{\mathbf{W}}(k) = A[\hat{\mathbf{W}}(k-1) + \mathbf{K}(k)\mathbf{E}(k)], \tag{11}$$

where

$$\hat{\mathbf{W}}(k) = [\hat{\mathbf{W}}_{0,0}^T(k), \cdots, \hat{\mathbf{W}}_{0,P-1}^T(k), \hat{\mathbf{W}}_{1,0}^T(k), \cdots, \hat{\mathbf{W}}_{1,P-1}^T(k)]^T \tag{12}$$

represents an estimate of the true weight vector $\mathbf{W}(k)$, $\mathbf{E}(k) = \mathbf{F}[\mathbf{0}_{1\times L}, e(kL), \cdots, e(kL+L-1)]^T$ is the frequency-domain error vector,

$$\hat{\mathbf{K}}(k) = [\hat{\mathbf{K}}_{0,0}^T(k), \cdots, \hat{\mathbf{K}}_{0,P-1}^T(k), \hat{\mathbf{K}}_{1,0}^T(k), \cdots, \hat{\mathbf{K}}_{1,P-1}^T(k)]^T \tag{13}$$

is the multi-channel Kalman gain comprising $2P$ Kalman gains with dimension $M \times M$, $\mathbf{P}(k)$ represents the multi-channel state-error covariance that includes $4P^2$ sub-matrix $\mathbf{P}_{i,j,p,q}(k), 0 \le i,j \le 1, 0 \le p,q \le P-1$ with dimension $M \times M$, and $\boldsymbol{\psi}_v(k) = E[\mathbf{V}(k)\mathbf{V}^H(k)]$ is the noise covariance matrix.

When the filter converges to the steady state, the error vector $\mathbf{E}(k)$ can be close to the noise vector. Therefore, the power spectral density (PSD) of the error signal can be used to estimate the noise PSD matrix

$$\hat{\boldsymbol{\psi}}_v(k) = \alpha\hat{\boldsymbol{\psi}}_v(k-1)+(1-\alpha)\text{diag}\left\{\mathbf{E}(k)\odot\mathbf{E}^H(k)\right\}, \tag{14}$$

where α is the smoothing factor.

Eq. (11) corresponds to the unconstrained FDAF algorithm. The convergence performance can be improved if we impose a constraint on the weight vector as follows [20]

$$\hat{\mathbf{W}}_{i,p}(k) = A[\hat{\mathbf{W}}_{i,p}(k-1) + \mathbf{G}_{10}\mathbf{K}_{i,p}(k)\mathbf{E}(k)], \tag{15}$$

where constraint matrix $\mathbf{G}_{10} = \mathbf{F}\mathbf{Q}_{10}^T\mathbf{Q}_{10}\mathbf{F}^{-1}$ forces the last half of the time-domain vector to zero with $\mathbf{Q}_{10} = [\ \mathbf{I}_L \quad \mathbf{0}_L\ ]$

3.2. Two simplified implementations

The direct evaluation of (10) involves an $M \times M$ matrices inversion, which is computationally intensive. We thus present two simplified versions based on a submatrix-diagonal form. We first assume that all the sub-matrix of $\mathbf{P}(k)$ are diagonal. We adopt the approximation [14]

$$\begin{aligned}&\mathbf{G}_{01}\mathbf{X}_{i,p}(k)\mathbf{P}_{i,j,p,q}(k)\mathbf{X}_{j,q}^H(k)\mathbf{G}_{01}^H\\ &\approx \tfrac{1}{2}\mathbf{X}_{i,p}(k)\mathbf{P}_{i,j,p,q}(k)\mathbf{X}_{j,q}^H(k).\end{aligned} \tag{16}$$

Using (16), we obtain

$$\begin{aligned}&\boldsymbol{\psi}(k) = \boldsymbol{\psi}_v(k)+\mathbf{X}(k)\mathbf{P}(k-1)\mathbf{X}^H(k) \approx \boldsymbol{\psi}_v(k)+\\ &\tfrac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{p=0}^{P-1}\sum_{q=0}^{P-1}\mathbf{X}_{i,p}(k)\mathbf{P}_{i,j,p,q}(k-1)\mathbf{X}_{j,q}^H(k).\end{aligned} \tag{17}$$

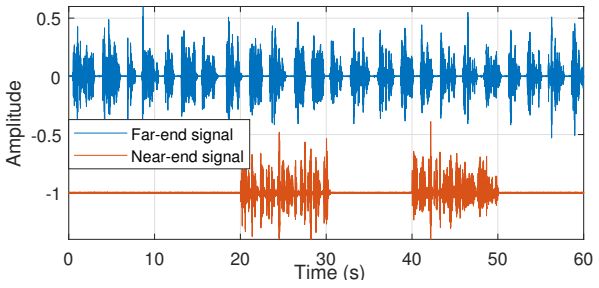

Figure 3: *Far-end source and near-end signal.*

Considering $\mathbf{G}_{01}\mathbf{X}_{i,p}(k) \approx \frac{1}{2}\mathbf{X}_{i,p}(k)$, we rewrite the Kalman gain $\mathbf{K}(k)$ as

$$\begin{aligned}\mathbf{K}(k) &= \mathbf{P}(k-1)\mathbf{X}^H(k)\boldsymbol{\psi}^{-1}(k)\\ &\approx \tfrac{1}{2}\mathbf{P}(k-1)\begin{bmatrix}\mathbf{X}_{0,0}^H(k)\boldsymbol{\psi}^{-1}(k)\\ \vdots\\ \mathbf{X}_{1,P-1}^H(k)\boldsymbol{\psi}^{-1}(k)\end{bmatrix},\end{aligned} \tag{18}$$

and we obtain $\mathbf{K}_{i,p}(k) \approx \frac{1}{2}\sum_{j=0}^{1}\sum_{q=0}^{P-1}\mathbf{P}_{i,j,p,q}(k-1)\mathbf{X}_{j,q}^H(k)\boldsymbol{\psi}^{-1}(k)$. Finally, we discuss the calculation of the state-error covariance matrix $\mathbf{P}(k)$. Since all sub-matrices of $\mathbf{P}(k)$ are diagonal, (9) can be rewritten as

$$\begin{aligned}&\mathbf{P}_{i,j,p,q}(k-1) \approx \boldsymbol{\psi}_{\Delta,i,j,p,q}(k)+A^2[\mathbf{P}_{i,j,p,q}(k-2)-\\ &\tfrac{1}{2}\mathbf{K}_{i,p}(k-1)\sum_{m=0}^{1}\sum_{l=0}^{P-1}\mathbf{P}_{m,j,l,q}(k-2)\mathbf{X}_{m,l}(k-1)].\end{aligned} \tag{19}$$

However, if all the sub-matrices $\mathbf{P}_{i,j,p,q}(k)$ are taken into account, we found that the PBFDKF algorithm is still computationally inefficient and also has a poor convergence behavior. We thus present two simplified versions. First, a natural choice is to only consider the covariance matrix on the main diagonal of $\mathbf{P}(k)$, i.e., $\mathbf{P}_{i,j,p,q}(k), i=j, p=q$, and ignore all other sub-matrices:

$$\mathbf{P}_{i,j,p,q}(k) = \mathbf{0}_M, i \ne j \parallel p \ne q. \tag{20}$$

Eq. (20) makes the proposed PBFDKF algorithm fully diagonal. This approach has already been used in [15]. The second simplified version is to consider the cross-correlation between the two sub-filters having the same partition index $\mathbf{P}_{i,j,p,q}(k), p=q, i \ne j$ and the auto-correlation terms $\mathbf{P}_{i,j,p,q}(k), p=q, i=j$, and then we have

$$\mathbf{P}_{i,j,p,q}(k) = \mathbf{0}_M, p \ne q. \tag{21}$$

We named the simplified versions in (20) and (21) as the variationally-diagonal PBFDKF (VD-PBFDKF) and the submatrix-diagonal PBFDKF (SD-PBFDKF), respectively.

We only focus on the two-channel case in this paper, but the proposed approach can be extended to the multichannel case straightforwardly.

4. Experiments and discussion

In this section, we conduct experiments to verify the performance of the proposed PBFDKF algorithm in the context of SEAC. Two far-end signals $x_1(n)$ and $x_2(n)$ are obtained by

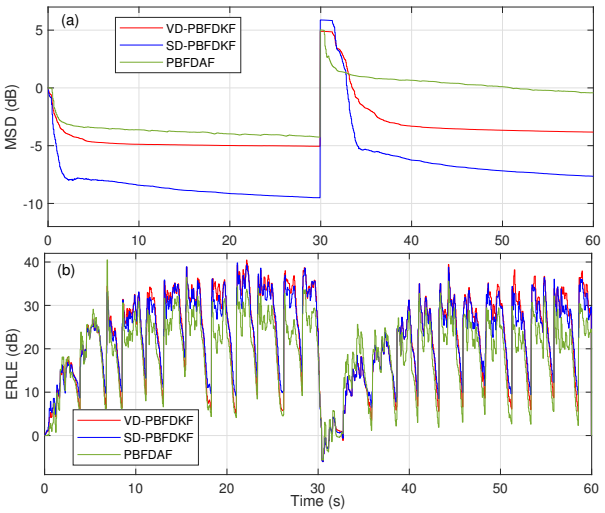

Figure 4: *Performance evaluation of standard stereophonic PBFDAF and two proposed algorithms in single-talk scenario, echo path is changed in the middle of the adaptation when* $t = 30$. *(a) MSD, (b) ERLE.*

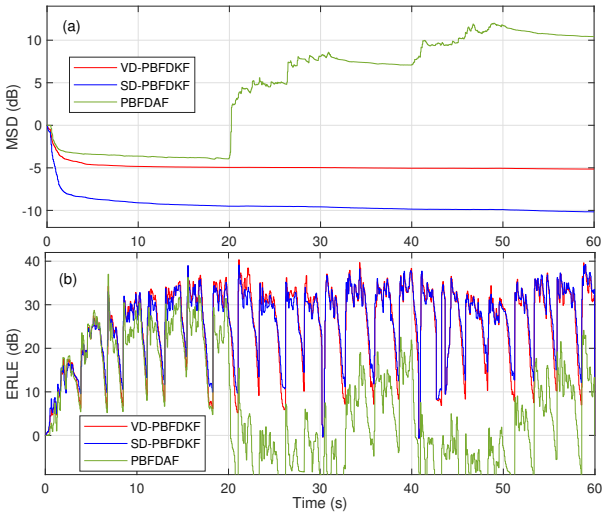

Figure 5: *Performance evaluation of standard stereophonic PBFDAF and two proposed algorithms in double-talk scenario, double-talk appears at 10-20 and 40-50 seconds, respectively. (a) MSD, (b) ERLE.*

convolving the far-end room transfer functions $g_1(n)$ and $g_2(n)$ with the far-end speaker $s(n)$, respectively. The sampling rate is 8000 Hz and the signal-to-noise ratio (SNR) at the far-end microphone is 40 dB. The length of the far-end room impulse responses $g_1(n)$ and $g_2(n)$ is 512. The stereo echo path in the near-end room is measured in an actual conference room and truncated to length $N = 2048$. We set the length of the sub-filter $L = 512$ and the number of blocks $P = 4$. The SNR at the near-end microphone is 30 dB. The far-end speaker source $s(n)$ and the near-end signal are shown in Fig. 3. We adopt the mean-square deviation, i.e., the system distance, and echo return loss enhancement (ERLE) to evaluate the performance of the algorithm, which are defined as, respectively,

$$\text{MSD} = 10\log_{10}[\frac{\left\|\mathbf{W}(k) - \hat{\mathbf{W}}(k)\right\|_2^2}{\|\mathbf{W}(k)\|_2^2}], \tag{22}$$

$$\text{ERLE} = 10\log_{10}\frac{E[|d(n) - v(n)|^2]}{E[|e(n) - v(n)|^2]}. \tag{23}$$

The MSD can evaluate how the estimated echo paths can approach to the true one, and the ERLE is used to evaluate the echo reduction performance. In the literature, the two criteria are indiscriminately used to evaluate the performance of AEC. The two-channel partitioned-block FDAF algorithm (PBFDAF) is also involved for comparison. To retain a high-quality speech, we did not employ any (non-linear) preprocessor here. We set the transition parameter $A = 0.9999$ and initialize state-error covariance as $\mathbf{P}_{i,j,p,q}(k) = \mathbf{I}_M$.

Fig. 4 shows the MSD and ERLE curves of the PBFDAF and the two proposed algorithms in single-talk scenario. The echo path is multiplied by -1 at the $t = 30$ s to simulate an abrupt echo-path change. When the transition parameter A is very close to one, the tracking performance of the proposed algorithm degrades due to the overestimation of the noise PSD. Several methods have been proposed to handle this problem. Here we use the shadow filter scheme [21]. Interested readers are referred to [10–12] for more details. It is apparent that the two proposed algorithms outperform the two-channel PBFDAF algorithm in terms of both MSD and ERLE.

Fig. 5 exhibits the comparison of the three algorithms in double-talk scenario. The near-end signal appears at 10-20 and 40-50 seconds, respectively, as depicted in Fig. 3. No double-talk detector is adopted here. The standard two-channel PBFDAF algorithm diverges rapidly for double-talk case, but the two proposed PBFDKF algorithms are very robust to near-end interference even without the double-talk detection. Observed from Figs. 4 and 5 the SD-PBFDKF outperforms the VD-PBFDKF in terms of MSD, but the VD-PBFDKF performs slightly better than the SD-PBFDKF in terms of ERLE. This is an interesting phenomenon that deserves further research. We also carried out experiments in a real acoustic scenario and obtained a similar result (not shown here for the space limitation).

5. Conclusions

This paper has proposed the PBFDKF algorithm to solve the SAEC problem. We derived the exact expression of the two-channel PBFDKF and then presented two low-complexity versions. Some practical problems, e.g., the noise covariance estimation and the tracking problem, were well considered. Experiments in real-time conference system showed that the proposed algorithms achieved satisfactory performance in both single-talk and double-talk scenarios. Interestingly, it was found that the two versions of PBFDKF behave differently in terms of MSD and ERLE, which has to be investigated more deeply in the future.

6. Acknowledgments

This work was supported by Youth Innovation Promotion Association of Chinese Academy of Sciences under Grant 2018027 and National Natural Science Foundation of China under Grant 11974376.

7. References

[1] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, 1998.

[2] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation-an overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 148–151, 1995.

[3] D. Morgan, J. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 686–696, 2001.

[4] L. Romoli, S. Cecchi, and F. Piazza, "Multichannel acoustic echo cancellation exploiting effective fundamental frequency estimation," *Speech Communication*, vol. 86, pp. 97–106, 2017.

[5] D.-Q. Nguyen, W.-S. Gan, and A. W. Khong, "Selective time-reversal block solution to the stereophonic acoustic echo cancellation problem," in *Proc. of 17nd European Signal Processing Conference*, Aug. 2009, pp. 1987–1991.

[6] J. Benesty, A. Gilloire, and Y. Grenier, "A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. L30–L35, 1999.

[7] S. Malik, J. Wung, J. Atkins, and D. Naik, "Double-talk robust multichannel acoustic echo cancellation using least-squares mimo adaptive filtering: Transversal, array, and lattice forms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4887–4902, 2020.

[8] Y. Gao, I. Liu, J. Z. C. Luo, and B. Li, "Independent echo path modeling for stereophonic acoustic echo cancellation," in *Proc. Interspeech*, 2020, pp. 3955–3958.

[9] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 373–376, 1990.

[10] 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, 2017.

[11] Z. Yan, F. Yang, and J. Yang, "Optimum step-size control for a variable step-size stereo acoustic echo canceller in the frequency domain," *Speech Communication*, vol. 124, pp. 21–27, 2020.

[12] F. Yang, G. Enzner, and J. Yang, "Statistical convergence analysis for optimal control of DFT-domain adaptive echo canceler," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1095–1106, 2017.

[13] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, 2006.

[14] S. Malik and G. Enzner, "Recursive bayesian control of multichannel acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 619–622, 2011.

[15] S. Malik and J. Benesty, "Variationally diagonalized multichannel state-space frequency-domain adaptive filtering for acoustic echo cancellation," in *Proc. IEEE ICASSP*, May 2013, pp. 595–599.

[16] M.-A. Jung, S. Elshamy, and T. Fingscheidt, "An automotive wideband stereo acoustic echo canceler using frequency-domain adaptive filtering," in *Proc. of 22nd European Signal Processing Conference*, Sep. 2014, pp. 1452–1456.

[17] S. Kühl, C. Antweiler, T. Hübschen, and P. Jax, "Kalman filter based stereo system identification with auto-and cross-decorrelation," in *Proc. HSCMA*, Mar. 2017, pp. 181–185.

[18] K. Sridhar, R. Cutler, A. Saabas, T. Parnamaa, H. Gamper, S. Braun, R. Aichner, and S. Srinivasan, "ICASSP 2021 acoustic echo cancellation challenge: Datasets and testing framework," in *Proc. IEEE ICASSP*, Jun. 2021, pp. 151–155.

[19] J.-M. Valin, S. Tenneti, K. Helwani, U. Isik, and A. Krishnaswamy, "Low-complexity, real-time joint neural echo control and speech enhancement based on percepnet," in *Proc. IEEE ICASSP*, Jun. 2021, pp. 7133–7137.

[20] S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[21] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceler with two echo path models," *IEEE Transactions on Communications*, vol. 25, no. 6, pp. 589–595, 1977.