2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC) | 979-8-3503-6185-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/IWAENC61483.2024.10694027

A DATA-REUSE SEMI-BLIND SOURCE SEPARATION APPROACH FOR NONLINEAR ACOUSTIC ECHO CANCELLATION

Yichen Yang[1,2], Xianrui Wang[1,2], Andreas Brendel[3], Wen Zhang[1], Jacob Benesty[4], Shoji Makino[2], and Jingdong Chen[1]

[1]Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an, China
[2]Graduate School of Information, Production and Systems, Waseda University, Kitakyushu, Japan
[3]Fraunhofer Institute for Integrated Circuits (IIS), D-91058 Erlangen, Germany
[4]INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada

ABSTRACT

Nonlinear acoustic echo cancellation (NAEC) is of significant importance in acoustic telecommunication. To improve NAEC performance in the double-talk case, semi-blind source separation-based NAEC (SBSS-NAEC) algorithms have been proposed. However, to deal with reverberation and loudspeaker nonlinearities, convolutive transfer function (CTF) models and power series expansions are employed, which significantly increase the number of free parameters and consequently lead to slow convergence speed and, hence, limited performance. In this paper, we introduce the data-reuse strategy, well-known in the adaptive filter literature, into an SBSS-NAEC framework and propose two algorithms: data-reuse iteration projection (DR-IP) and data-reuse element-wise iterative source steering (DR-EISS). Several simulations demonstrate the superiority of the proposed methods, especially the tracking capability when the impulse response changes.

Index Terms— Nonlinear acoustic echo cancellation, semi-blind source separation, data reuse, iterative projection, element-wise iterative source steering

1. INTRODUCTION

Nonlinear acoustic echo cancellation (NAEC) has been a subject of extensive research over the past few decades [1–3]. Generally, the loudspeaker nonlinearities are approximated with a power series expansion, and the acoustic impulse response (AIR) from the loudspeaker to the microphone is modeled as a finite impulse response (FIR) filter. Typically, adaptive filters are used to estimate corresponding expansion weights and FIR coefficients [4–10]. However, such methods still suffer from several major issues including: 1) low convergence rates and bad tracking capabilities, and 2) degradation of convergence rate and performance during activity of the near-end signal, i.e., in the double talk situation.

To maintain the effectiveness of echo cancellation in double-talk scenarios, a straightforward way is to incorporate a double-talk detector into the original NAEC methods, resulting in the suspension of filter updates during double-talk situations. Therefore, echo cancellation quality depends on the detection accuracy of double-talk situations. Furthermore, double-talk detection introduces an additional algorithmic delay, impacting real communication system performance. On the other hand, semi-blind source separation (SBSS)-based NAEC, has been proposed as an alternative technique for eliminating the echo during double-talk situations without the need of explicit double-talk detection [11, 12] from the perspective of blind source separation (BSS) [13–18]. In strongly reverberant environments, convolutive transfer function (CTF)-based models [19] are often used to maintain performance for BSS [20, 21] and SBSS-based NAEC tasks [22–24].

By reusing the same data (i.e., segments of the input and reference signals) several times for computing the filter updates, the data-reuse (DR) scheme has been proposed for adaptive filter-based NAEC algorithms such as DR LMS [25, 26], DR NLMS [27–29], and DR RLS [30–32] approaches. Here, a significant improvement of the convergence rates and tracking capabilities has been shown by using this strategy. Hence, in this paper, we incorporate the DR scheme into the SBSS-based NAEC framework and present two novel algorithms, which could be regarded as an extension of our previous work [23]. By reusing the same data when updating each frame, the convergence rate is improved significantly, while simultaneously enhancing the overall performance. Simulations demonstrate that the proposed method significantly improves cancellation performance relative to baselines in double-talk situations.

2. SIGNAL MODEL AND PROBLEM FORMULATION

Considering a full-duplex speech communication scenario, in the time domain, the signal observed by the microphone $y(t)$ can be expressed as

$$\begin{aligned} y(t) &= s(t) + v(t) \\ &= s(t) + a(t) * f\left[x(t)\right], \end{aligned} \tag{1}$$

where t is the time index, $s(t)$ and $x(t)$ are the near-end and far-end signals, respectively, $v(t) = a(t) * f\left[x(t)\right]$ is the acoustic echo, $a(t)$ is the AIR, $f[\cdot]$ denotes the distortion caused by the loudspeaker, which includes both linear and nonlinear effects, and $*$ denotes the linear convolution.

To overcome the difficulty of estimating a nonlinear system, in practice, the nonlinear effects caused by the loudspeaker is commonly approximated by a linear combination of the power series of the far-end signal [33, 34], like the odd power series expansion

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2021ZD0201502 and in part by the NSFC Grants: 62171373, 61831019, 62192713 and 62271401, and in part by the China Scholarship Council (CSC).

[22, 35], which can be expressed as

$$f\left[x(t)\right] = \sum_{p=0}^{P-1} c_p x^{2p+1}(t), \tag{2}$$

where p and P are the index of the expansion coefficient c_p and the expansion order, respectively. By inserting the representation (2) into the signal model (1), the observed signal can be denoted as

$$y(t) = s(t) + a(t) * \left(\sum_{p=0}^{P-1} c_p x^{2p+1}(t) \right). \tag{3}$$

By using the CTF-based model, the observed signal can be rewritten in the STFT domain as

$$\begin{aligned} Y_{i,j} &= S_{i,j} + \sum_{l=0}^{L-1} \sum_{p=0}^{P-1} c_p A_{i,j,l} X_{p,i,j-l} \\ &= S_{i,j} + \sum_{l=0}^{L-1} \sum_{p=0}^{P-1} A'_{p,i,j,l} X_{p,i,j-l}, \end{aligned} \tag{4}$$

where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$ are indexes of frequency bins and time frames, respectively. Hereby, I denotes the number of frequency bins and J is the number of time frames, L is the CTF filter length and $A_{i,j,l}$ represents the acoustic transfer function (ATF). Furthermore, we introduced $A'_{p,i,j,l} = c_p A_{i,j,l}$ as the merged ATF, and $Y_{i,j}$, $S_{i,j}$, and $X_{p,i,j-l}$ as the STFT-domain representations of $y(t)$, $s(t)$, and $x^{2p+1}(t)$, respectively. The mixing model (4) can be compactly written in matrix form as

$$\widetilde{\mathbf{y}}_{i,j} = \mathbf{H}'_{i,j} \widetilde{\mathbf{s}}_{i,j}, \tag{5}$$

with

$$\mathbf{x}_{i,j-l} = \left[X_{0,i,j-l} \;\; X_{1,i,j-l} \;\cdots\; X_{P-1,i,j-l}\right]^T \in \mathbb{C}^{P\times 1}, \tag{6}$$

$$\widetilde{\mathbf{s}}_{i,j} = \left[S_{i,j} \;\; \mathbf{x}_{i,j}^T \;\; \mathbf{x}_{i,j-1}^T \;\cdots\; \mathbf{x}_{i,j-L+1}^T\right]^T \in \mathbb{C}^{(PL+1)\times 1}, \tag{7}$$

$$\widetilde{\mathbf{y}}_{i,j} = \left[Y_{i,j} \;\; \mathbf{x}_{i,j}^T \;\; \mathbf{x}_{i,j-1}^T \;\cdots\; \mathbf{x}_{i,j-L+1}^T\right]^T \in \mathbb{C}^{(PL+1)\times 1}, \tag{8}$$

$$\mathbf{a}'_{i,j,l} = \left[A'_{0,i,j,l} \;\; A'_{1,i,j,l} \;\cdots\; A'_{P-1,i,j,l}\right]^T \in \mathbb{C}^{P\times 1}, \tag{9}$$

$$\widetilde{\mathbf{a}}'_{i,j} = \left[(\mathbf{a}'_{i,j,0})^T \;\; (\mathbf{a}'_{i,j,1})^T \;\cdots\; (\mathbf{a}'_{i,j,L-1})^T\right]^T \in \mathbb{C}^{PL\times 1}, \tag{10}$$

$$\mathbf{H}'_{i,j} = \begin{bmatrix} 1 & (\widetilde{\mathbf{a}}'_{i,j})^T \\ \mathbf{0}_{PL\times 1} & \mathbf{I}_{PL} \end{bmatrix} \in \mathbb{C}^{(PL+1)\times(PL+1)}, \tag{11}$$

where $(\cdot)^T$ represents the transpose operator, $\mathbf{I}_{PL}$ is the identity matrix of size $PL \times PL$, and $\mathbf{H}'_{i,j}$ is the mixing matrix producing the observed signal. To extract the near-end signal $S_{i,j}$ from the observations and to suppress the echo, we take an BSS perspective and calculate the separated signals $\hat{\mathbf{s}}_{i,j}$ via the demixing matrix $\mathbf{W}_{i,j}$, which we choose here to be the inverse of the mixing matrix:

$$\mathbf{W}_{i,j} = \begin{bmatrix} 1 & -(\widetilde{\mathbf{a}}'_{i,j})^T \\ \mathbf{0}_{PL\times 1} & \mathbf{I}_{PL} \end{bmatrix} \in \mathbb{C}^{(PL+1)\times(PL+1)}, \tag{12}$$

by

$$\begin{aligned} \hat{\mathbf{s}}_{i,j} &= \mathbf{W}_{i,j} \widetilde{\mathbf{y}}_{i,j} \\ &= \left[\hat{S}_{i,j} \;\; \mathbf{x}_{i,j}^T \;\; \mathbf{x}_{i,j-1}^T \;\cdots\; \mathbf{x}_{i,j-L+1}^T\right]^T, \end{aligned} \tag{13}$$

where $\hat{S}_{i,j}$ is the estimate of the near-end signal $S_{i,j}$.

3. PROPOSED METHOD

3.1. Probabilistic Model

To estimate the demixing matrix and to obtain the near-end signal estimate, we assume that the near-end signal follows a super Gaussian distribution, which can be expressed as

$$p(\underline{\mathbf{s}}_j) \propto \exp\left[-\left(\frac{\left\|\underline{\mathbf{s}}_j\right\|_2}{\gamma} \right)^{\beta} \right], \tag{14}$$

where $\underline{\mathbf{s}}_j = \left[S_{1,j} \; S_{2,j} \;\cdots\; S_{I,j}\right]^T$ is the vector of the near-end signal, $0 < \beta < 2$ and $\gamma > 0$ are the shape and scale parameters, respectively, and $\|\cdot\|_2$ represents the ℓ_2 norm. Based on an I.I.D. assumption, the following negative log-likelihood cost function is formulated

$$\begin{aligned} \mathcal{L}_j = &- \frac{1}{\sum_{j'=1}^{j} \alpha^{j-j'}} \sum_{j'=1}^{j} \alpha^{j-j'} \log p(\underline{\mathbf{s}}_{j'}) \\ &- 2 \sum_{i=1}^{I} \log \left|\det \mathbf{W}_{i,j}\right|, \end{aligned} \tag{15}$$

where $0 < \alpha < 1$ is the forgetting factor. Similar to auxiliary function-based IVA (AuxIVA), we derive a surrogate of the loss (15) by means of the majorize-minimization (MM) algorithm

$$\mathcal{L}_j^+ = \sum_{i=1}^{I} \mathbf{w}_{i,j}^H \mathbf{V}_{i,j} \mathbf{w}_{i,j} - 2 \sum_{i=1}^{I} \log \left|\det \mathbf{W}_{i,j}\right|, \tag{16}$$

where $(\cdot)^H$ represents the conjugate-transpose operator, $\mathbf{w}_{i,j}^H = \left[1 \quad -(\widetilde{\mathbf{a}}'_{i,j})^T\right]$ is the first row of the demixing matrix $\mathbf{W}_{i,j}$. In (16), we introduced the auxiliary variable $\mathbf{V}_{i,j}$, which can be expressed as

$$\mathbf{V}_{i,j} = \alpha \mathbf{V}_{i,j-1} + (1-\alpha) \varphi(r_j) \widetilde{\mathbf{y}}_{i,j} \widetilde{\mathbf{y}}_{i,j}^H, \tag{17}$$

where

$$\varphi(r_j) = r_j^{\beta-2}, \tag{18}$$

$$r_j = \sqrt{\sum_{i=1}^{I} \left|\mathbf{w}_{i,j-1}^H \widetilde{\mathbf{y}}_{i,j}\right|^2}. \tag{19}$$

To optimize the cost function (16), classically, the update rules of AuxIVA-based SBSS AEC update the demixing filter only once for each time frame, i.e., $\widetilde{\mathbf{y}}_{i,j}$ in (17) [12, 23]. As the original cost function (15) is non-convex, a single MM iteration may not yield optimal results. Consequently, in this paper, we utilize a data-reuse strategy to address this challenge, and two algorithms have been proposed: the data-reuse iteration projection (DR-IP) and data-reuse element-wise iterative source steering (DR-EISS).

3.2. Data-Reuse IP-Based Optimization

By assuming that the demixing filter $\mathbf{w}_{i,j}$ is updated N times using the same data $\widetilde{\mathbf{y}}_{i,j}$, the loss in (16) and update (17) of the auxiliary variable $\mathbf{V}_{i,j}$ can be rewritten as

$$\mathcal{L}^{+}_{j;n} = \sum_{i=1}^{I} \mathbf{w}^H_{i,j;n}\mathbf{V}_{i,j;n}\mathbf{w}_{i,j;n} - 2\sum_{i=1}^{I} \log|\det \mathbf{W}_{i,j;n}| \tag{20}$$

and

$$\mathbf{V}_{i,j;n} = \alpha \mathbf{V}_{i,j;n-1} + (1-\alpha)\varphi(r_{j;n})\widetilde{\mathbf{y}}_{i,j}\widetilde{\mathbf{y}}^H_{i,j}, \tag{21}$$

with

$$\varphi(r_{j;n}) = r^{\beta-2}_{j;n}, \tag{22}$$

$$r_{j;n} = \sqrt{\sum_{i=1}^{I} \left|\mathbf{w}^H_{i,j;n-1}\widetilde{\mathbf{y}}_{i,j}\right|^2}, \tag{23}$$

where $n = 1, 2, \ldots, N$ is the iteration index for each time frame j and $\mathbf{V}_{i,j;n}$ is the auxiliary variable after reusing the data of the jth frame n times. Note that for the first iteration $n = 1$, the auxiliary variable and the demixing filter corresponding to the jth frame can be initialized by the result of the Nth iteration of the previous frame, i.e., $\mathbf{V}_{i,j;0} = \mathbf{V}_{i,j-1;N}$ and $\mathbf{w}_{i,j;0} = \mathbf{w}_{i,j-1;N}$. By equating the derivative of (20) with respect to $\mathbf{w}_{i,j;n}$ to 0, the demixing filter $\mathbf{w}_{i,j;n}$ can be updated by regular IP update rules as [15]

$$\begin{aligned}\mathbf{w}_{i,j;n} &\leftarrow \left(\mathbf{W}_{i,j;n}\mathbf{V}_{i,j;n}\right)^{-1}\mathbf{e}_1 \\ &= \mathbf{V}^{-1}_{i,j;n}\mathbf{e}_1,\end{aligned} \tag{24}$$

where $\mathbf{e}_1$ is the unit vector with the first element being 1 and all others being 0 and $\mathbf{w}_{i,j;n}$ is the demixing filter of the nth iteration at the jth frame. Furthermore, to satisfy the structure of $\mathbf{w}_{i,j;n}$ (see first row of (12)), $\mathbf{w}_{i,j;n}$ is further normalized as [12]

$$\mathbf{w}_{i,j;n} \leftarrow \mathbf{w}_{i,j;n}/w_{i,j,1;n}, \tag{25}$$

to ensure that the first element of $\mathbf{w}_{i,j;n}$ is fixed to one, where $w_{i,j,1;n}$ is the first element of $\mathbf{w}_{i,j;n}$.

3.3. Data-Reuse EISS-Based Optimization

In order to reduce the computational cost of IP-based optimization, recently, a more efficient updating rule based on the element-wise iterative source steering (EISS) has been proposed [23], and the updating rules with data-reuse strategy can be given as

$$w_{i,j,k;n} = \begin{cases} w_{i,j,k;n-1} - u_{i,j,k;n}, & \text{if } k=1, \\ w_{i,j,k;n-1} - u_{i,j,1;n}w_{i,j,k;n-1} - u_{i,j,k;n}, & \text{else.} \end{cases} \tag{26}$$

Here, $w_{i,j,k;n}$ with $k = 1, 2, \ldots, PL+1$ is the kth element of $\mathbf{w}_{i,j;n}$ and $u_{i,j,k;n}$ is the parameter to be estimated. With the updating rules in (26), the auxiliary function in (20) can be rewritten as

$$\begin{aligned}\mathcal{L}^{+}_{i,j;n} =& \left(\mathbf{w}_{i,j;n-1} - \mathbf{d}_{i,j;n}\right)^H \mathbf{V}_{i,j;n}\left(\mathbf{w}_{i,j;n-1} - \mathbf{d}_{i,j;n}\right) \\ &- 2\log|1 - u_{i,j,1;n}|,\end{aligned} \tag{27}$$

where

$$\begin{aligned}\mathbf{d}_{i,j;n} = [&u_{i,j,1;n} \;\; u_{i,j,1;n}w_{i,j,2;n-1} + u_{i,j,2;n} \\ &\cdots \;\; u_{i,j,1;n}w_{i,j,PL+1;n-1} + u_{i,j,PL+1;n}]^H.\end{aligned} \tag{28}$$

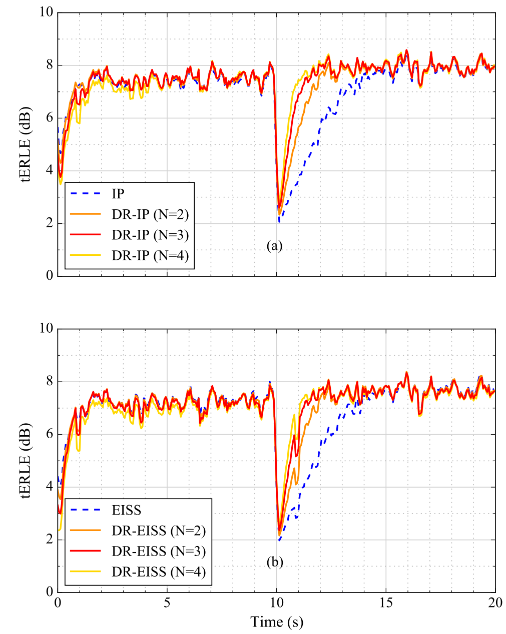

Fig. 1. Comparison of average tERLE (dB) during double-talk with different optimization approaches for the demixing filter: (a) IP, (b) EISS. The far-end signals are a white Gaussian noise filtered by an autoregressive model and the environment changes after 10 seconds.

Finally, the $u_{i,j,k;n}$ can be obtained by setting the derivative of (27) with respect to $u^*_{i,j,k;n}$ to 0, which results in

$$u_{i,j,k;n} = \begin{cases} 1 - \left(\mathbf{w}^H_{i,j;n-1}\mathbf{V}_{i,j;n}\mathbf{w}_{i,j;n-1}\right)^{-\frac{1}{2}}, & \text{if } k=1, \\ \mathbf{w}^H_{i,j;n-1}\mathbf{v}_{i,j,k;n}/V_{i,j,k,k;n}, & \text{else,} \end{cases} \tag{29}$$

where $\mathbf{v}_{i,j,k;n}$ is the kth column of $\mathbf{V}_{i,j;n}$ and $V_{i,j,k,k;n}$ is the kth element of $\mathbf{v}_{i,j,k;n}$. By using the $u_{i,j,k;n}$ from (29), $\mathbf{w}_{i,j;n}$ can be updated from (26) sequentially followed by normalization with (25). Since the EISS-based updating rules do not rely on matrix inversion, the algorithmic complexity is decreased remarkably.

4. EXPERIMENTAL EVALUATION

4.1. Experimental Setup

The performance of the proposed algorithms is verified through simulations in this section. To simulate a double-talk scenario in realistic acoustic environments, a GPU-accelerate image method [36], referred to as gpuRIR [37], is applied to simulate the room impulse responses (RIRs) between the two sources and the microphone with a reverberation time of approximate 300 ms. For the near-end signal, i.e., $s(t)$, 20 speech utterances are randomly picked from the CMU ARCTIC dataset [38]. The energy-based voice activity detection (VAD) used in [39] is applied to remove the silence chunk in each speech utterance and each utterance is then trimmed to 10 seconds. For the far-end signal, i.e., $x(t)$, we consider two different signal classes to highlight different aspects of the proposed methods: white Gaussian noise filtered by an autoregressive process and the speech signal, which is also randomly selected from the CMU

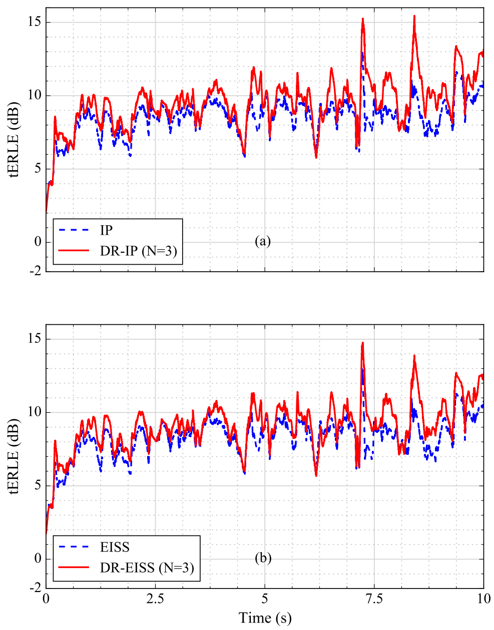

Fig. 2. Comparison of tERLE (dB) during double-talk with different optimization approaches for the demixing filter: (a) IP, (b) EISS. The position of the near-end speaker remains stable.

ARCTIC dataset. Hard clipping is used to simulate loudspeaker distortions, which can be expressed as

$$f\left[x(t)\right] = \begin{cases} -x_{\text{thr}}, & \text{if } x(t) < -x_{\text{thr}}, \\ x(t), & \text{if } |x(t)| \leq x_{\text{thr}}, \\ x_{\text{thr}}, & \text{if } x(t) > x_{\text{thr}}, \end{cases} \tag{30}$$

where x_{thr} is the threshold of the clipping and is set to 0.2 times of the max value of $|x(t)|$, which is consistent with that in the paper [22].

Finally, the echos are simulated by convolving the distorted far-end signal with the RIRs. The observed signals are generated by adding the near-end signal and the echo at an input signal-to-echo ratio (SER) of 0 dB. The true echo return loss enhancement (tERLE) [35] is used as the evaluation measure. All signals are sampled at 16 kHz. The STFT is implemented with a von Hann window of 16 ms length, and a window shift of 4 ms. The expansion order P is set to 3, the length of the CTF filter L is set to 5, the forgetting factor α is set to 0.998, and the shape parameter β is set to 0.4. The demixing matrix is initialized as an identity matrix $\mathbf{I}_{PL+1}$.

The experimental results of the proposed methods (referred to as DR SBSS-CTF-AuxIVA-IP and DR SBSS-CTF-AuxIVA-EISS) is compared with the SBSS-CTF-AuxIVA-IP in [22] and the SBSS-CTF-AuxIVA-EISS in [23].

4.2. Results and Discussion

The convergence rate of the presented algorithms and the baselines relying on different optimization approaches for the demixing filter are presented in Fig. 1. In this experiment, the far-end signals are white Gaussian noise filtered by an autoregressive process. By updating the demixing filter with DR in each frame, the proposed method shows fast and stable convergence. Furthermore, when the RIRs changes after 10 seconds, the proposed method also shows a good tracking ability for both optimization methods.

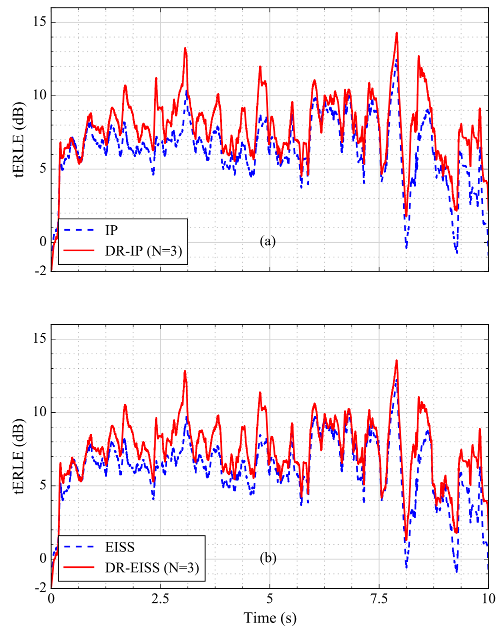

Fig. 3. Comparison of tERLE (dB) during double-talk with different optimization approaches for the demixing filter: (a) IP, (b) EISS. The impulse response between the loudspeaker and microphone changes.

Scenarios	IP	EISS	**DR-IP**	**DR-EISS**
Stable	8.50	8.28	9.54	9.13
Changing	6.33	6.11	7.97	7.56

Table 1. The average tERLE (dB) of the present methods.

Fig. 2 and Fig. 3 show the AEC performance when the impulse response is stable and changing scenarios, respectively, where the speech signal is utilized as the far-end signal. The average tERLE under these two scenarios is shown in Table 1. Compared to the baselines, the proposed DR-based method shows a significant improvement in tERLE, which indicates that with the DR scheme, the SBSS-based AEC system can efficiently speed up convergence and enhances the performance. Furthermore, when the near-end speaker moves (impulse response changes), the performance improvement achieved with the DR scheme is more significant compared to that in a stable environment. This highlights the DR scheme's strong tracking ability, especially in impulse response varying scenarios.

5. CONCLUSIONS

In this paper, we incorporated the DR scheme into an SBSS-based AEC framework relying on CTF-AuxIVA for estimating the demixing filter. We derived IP-based and EISS-based optimization methods and showed that the presented algorithm works well in terms of both convergence rate (tracking ability) and overall AEC performance. The presented simulation results demonstrate the superiority of the proposed method over the compared baseline methods.

6. REFERENCES

[1] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE ICASSP*, 1988, pp. 2570–2573.

[2] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in network and acoustic echo cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[3] Y. Huang, J. Chen, and J. Benesty, "Immersive audio schemes," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 20–32, Jan. 2011.

[4] E. Ferrara, "Fast implementations of LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 474–475, Aug. 1980.

[5] G. Long, F. Ling, and J. G. Proakis, "The LMS algorithm with delayed coefficient adaptation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 9, pp. 1397–1405, Sept. 1989.

[6] J.-I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, Jun. 1967.

[7] C. Paleologu, J. Benesty, and S. Ciochină, "An improved proportionate NLMS algorithm based on the l_0 norm," in *Proc. IEEE ICASSP*, 2010, pp. 309–312.

[8] V Panuska, "An adaptive recursive-least-squares identification algorithm," in *Proc. IEEE ICASSP*, 1969, pp. 65–65.

[9] C. Paleologu, J. Benesty, and S. Ciochină, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, Oct. 2008.

[10] W. Wang, Y. Na, Z. Liu, B. Tian, and Q. Fu, "Weighted recursive least square filter and neural network based residual echo suppression for the AEC-challenge," in *Proc. IEEE ICASSP*, 2021, pp. 141–145.

[11] F. Nesta, T. S. Wada, and B.-H. Juang, "Batch-online semi-blind source separation applied to multi-channel acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 583–599, Mar. 2011.

[12] G. Cheng, L. Liao, H. Chen, and J. Lu, "Semi-blind source separation for nonlinear acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 28, pp. 474–478, Feb. 2021.

[13] S. Makino, T.-W. Lee, and H. Sawada, *Blind speech separation*. Dordrecht, Netherlands: Springer, 2007.

[14] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. Hoboken, NJ, USA: John Wiley & Sons, 2018.

[15] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE WASPAA*, 2011, pp. 189–192.

[16] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *Proc. IEEE ICASSP*, 2020, pp. 236–240.

[17] Y. Yang, X. Wang, A. Brendel, W. Zhang, W. Kellermann, and J. Chen, "Geometrically constrained source extraction and dereverberation based on joint optimization," in *Proc. EUSIPCO*, 2023, pp. 41–45.

[18] K. Mo, X. Wang, Y. Yang, T. Ueda, S. Makino, and J. Chen, "On joint dereverberation and source separation with geometrical constraints and iterative source steering," in *Proc. APSIPA ASC*, 2023, pp. 1138–1142.

[19] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 546–555, May 2009.

[20] T. Wang, F. Yang, and J. Yang, "Convolutive transfer function-based multichannel nonnegative matrix factorization for overdetermined blind source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 802–815, Jan. 2022.

[21] X. Wang, A. Brendel, G. Huang, Y. Yang, W. Kellermann, and J. Chen, "Spatially informed independent vector analysis for source extraction based on the convolutive transfer function model," in *Proc. IEEE ICASSP*, 2023, pp. 1–5.

[22] G. Cheng, L. Liao, K. Chen, Y. Hu, C. Zhu, and J. Lu, "Semi-blind source separation using convolutive transfer function for nonlinear acoustic echo cancellation," *J. Acoust. Soc. Am.*, vol. 153, no. 1, pp. 88–95, Jan. 2023.

[23] K. Lu, X. Wang, T. Ueda, S. Makino, and J. Chen, "A computationally efficient semi-blind source separation approach for nonlinear echo cancellation based on an element-wise iterative source steering," in *Proc. IEEE ICASSP*, 2024, pp. 756–760.

[24] X. Wang, Y. Yang, A. Brendel, T. Ueda, S. Makino, J. Benesty, W. Kellermann, and J. Chen, "On semi-blind source separation-based approaches to nonlinear echo cancellation based on bilinear alternating optimization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 2973–2987, May 2024.

[25] S. Shaffer and C. S. Williams, "Comparison of LMS, alpha LMS, and data reusing LMS algorithms," in *Proc. Conf. Rec. 17th Asilomar Conf. Circuits, Syst. Comput.*, 1983, pp. 260–264.

[26] S. Roy and J. J. Shynk, "Analysis of the data-reusing LMS algorithm," in *Proc. 32nd Midwest Symp. Circuits Syst.*, 1989, pp. 1127–1130.

[27] B. A. Schnaufer, W. K. Jenkins, "New data-reusing LMS algorithms for improved convergence," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 1584–1588.

[28] J. Benesty and T. Gänsler, "On data-reuse adaptive algorithms," in *Proc. IEEE IWAENC*, 2003, pp. 31–34.

[29] C. Paleologu and J. Benesty, "A practical data-reuse adaptive algorithm for acoustic echo cancellation," in *Proc. EUSIPCO*, 2012, pp. 2010–2014.

[30] C. Paleologu, J. Benesty, and S. Ciochină, "Data-reuse recursive least-squares algorithms," *IEEE Signal Process. Lett.*, vol. 29, pp. 752–756, Feb. 2022.

[31] L.-M. Dogariu, C. Paleologu, J. Benesty, and S. Ciochină, "On the performance of a data-reuse fast RLS algorithm for acoustic echo cancellation," in *Proc. IEEE BlackSeaCom*, 2022, pp. 135–140.

[32] W. Gao, J. Chen, and C. Richard, "Theoretical analysis of the performance of the data-reuse RLS algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, pp. 1–1, Aug. 2023.

[33] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE ICASSP*, 2005, pp. 105–108.

[34] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Process.*, vol. 86, pp. 1168–1181, Nov. 2006.

[35] S. Malik and G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 7, pp. 2065–2079, Dec. 2012.

[36] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[37] D. Diaz-Guerra, A. Miguel, and J. R. Beltran, "gpuRIR: A python library for room impulse response simulation with GPU acceleration," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5653–5671, Feb. 2021.

[38] J. Kominek and A. W. Black, "CMU arctic databases for speech synthesis," *Carnegie Mellon Univ., Tech. Rep. CMU-LTI-03-177*, 2003.

[39] Y. Luo and J. Yu, "Music source separation with band-split RNN," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 1893–1901, May 2023.