On Semi-Blind Source Separation-Based Approaches to Nonlinear Echo Cancellation Based on Bilinear Alternating Optimization

Xianrui Wang, *Graduate Student Member, IEEE*, Yichen Yang, *Graduate Student Member, IEEE*, Andreas Brendel, *Member, IEEE*, Tetsuya Ueda, *Student Member, IEEE*, Shoji Makino, *Life Fellow, IEEE*, Jacob Benesty, Walter Kellermann, and Jingdong Chen, *Fellow, IEEE*

***Abstract*—Acoustic echo cancellation (AEC) is a crucial task in full duplex communications. As conventional linear filtering approaches are ineffective to deal with double-talk, various semi-blind source separation (SBSS)-based AEC algorithms are deceived, most of which are formulated and implemented in the frequency domain based on the multiplicative transfer function (MTF) model for computational efficiency. To avoid large latency and in order to deal with loudspeaker nonlinearities, the convolutive transfer function (CTF) model and odd power series expansion are leveraged, which are employed by numerous SBSS-based nonlinear AEC (SBSS-NAEC) algorithms. Conventional SBSS-NAEC methods estimate the series expansion coefficients and the CTF filter simultaneously making the number of free parameters to estimate large. Hence, the corresponding algorithms are computationally expensive and are difficult to optimize. In this work, we propose to decouple the series expansion coefficients and the CTF filters into a bilinear form and present a bilinear alternating optimization framework for estimating the model parameters. An alternating iterative projection (AIP) algorithm and an alternating element-wise iterative source steering (AEISS) algorithm are proposed. As the bilinear representation consists of less parameters compared to the conventional methods, the proposed algorithms not only improve the AEC performance but also reduce the computational complexity, which is validated by comprehensive simulations and experiments.**

***Index Terms*—Semi-blind source separation, nonlinear acoustic echo cancellation, odd power series expansion, convolutive transfer function model, bilinear, alternating optimization.**

Manuscript received 4 February 2024; revised 14 May 2024; accepted 20 May 2024. Date of publication 30 May 2024; date of current version 10 June 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021ZD0201502 and in part by the Major Program of National Science Foundation of China (NSFC) under Grant 62192713. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jan Skoglund. *(Corresponding author: Jingdong Chen.)*

Xianrui Wang and Yichen Yang are with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an 710071, China, and also with the Waseda University, Tokyo 169-8050, Japan.

Andreas Brendel is with the Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS in Erlangen, 91058 Erlangen, Germany.

Tetsuya Ueda and Shoji Makino are with the Graduate School of Information, Production and Systems, Waseda University, Kitakyushu 808-0135, Japan.

Jacob Benesty is with the INRS-EMT, University of Quebec, Montreal, Quebec 3C 3P8, Canada.

Walter Kellermann is with the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, 91054 Erlangen, Germany.

Jingdong Chen is with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an 710071, China (e-mail: jingdongchen@ieee.org).

Digital Object Identifier 10.1109/TASLP.2024.3407676

I. Introduction

Acoustic echoes, which are caused by coupling between the loudspeakers and microphones, are detrimental to full duplex voice communication [1], [2], [3]. Consequently, acoustic echo cancellation (AEC), a process to estimate and eliminate echoes, has to be used in full duplex voice communication systems [1], [2], [3], [4], [5], [6]. Generally, AEC models the acoustic impulse response (AIR) from the loudspeaker to the microphone as a finite impulse response (FIR) filter. The problem is then transformed into one of FIR channel identification [3], [7], [7], [8], [9], [10], [11], [12]. Since AIRs are generally time-varying, adaptive filters have to be used and many adaptive algorithms have been developed over the last few decades, such as the least-mean-square (LMS) [13], [14], [15], [16], normalized LMS (NLMS) [17], [18], [19], [20], recursive least-squares (RLS) [21], [22], [23], [24], and Kalman filters [25], [26], [27]. Those algorithms can achieve good AEC performance in the single-talk scenario, where there is no near-end speech; but their performance often degrades significantly in the presence of double-talk, where both the far- and near-end speech signals coexist. The traditional way to deal with this issue is through using a double-talk detector (DTD) [28], [29] and the adaptation process stops whenever double-talk is detected. While it has been proved to be a viable way to handle double-talk, this approach is unable to track the dynamic acoustic system when the adaptation is stalled and, as a result, often leads to severe performance degradation during double-talk. To address this issue, some Kalman filter-like algorithms were developed, which model both the near-end and echo signals as uncorrelated complex Gaussian distributed processes [30], [31]. While improvement is observed, the performance of these algorithms remains notably inadequate and fails to meet the demands of practical systems.

Alternatively, the double-talk issue can be addressed from the perspective of blind source separation (BSS) [32], [33], [34], and several semi-blind source separation-based AEC (SBSS-AEC) algorithms have been developed [35], [36], [37], [38], which are formulated in the frequency domain and can be implemented efficiently thanks to the fast Fourier transform (FFT). Early such algorithms adopted the so-called multiplicative transfer function (MTF) model in which the time-domain convolution is represented by the frequency-domain bin-wise multiplication [35],

[38]. To achieve good performance, the analysis window for short-time Fourier transform (STFT) in this model has to be long enough to cover the effective part of the AIR (i.e, the components that significantly contribute to the echo). This, however, leads to large algorithmic latency, particularly when the system operates in large rooms or enclosures with high level of reverberation [39]. To circumvent this issue, the convolutive transfer function (CTF) model was developed, which represents the time-domain convolution with a frequency-domain convolution using signal components from multiple consecutive frames [40], [41], [42], [43] and is structurally equivalent to frequency subband filtering [44], [45]. This framework offers flexibility to choose shorter analysis window, which enables to reduce the algorithmic latency. The CTF filter estimation can be formulated as a simplified hybrid exact-approximate diagonalization (HEAD) problem [46] based on the assumption that the far- and near-end signals are mutually independent, which has been well studied in the literature of BSS [47], [48], [49], [50], [51], [52], [53], [54]. Iterative projection (IP) [47] is one of the most widely-used algorithms to solve this problem for SBSS-AEC [55], [56]. However, since it involves matrix inversion, the IP algorithm is computationally very expensive. With the inspiration of the iterative source steering (ISS) principle [52], a more computationally efficient algorithm was developed in [57], which is called the element-wise iterative source steering (EISS). As shown in [56], [57], both the IP and EISS algorithms can achieve significant performance improvement in comparison with the aforementioned Kalman filter for AEC in double-talk scenarios [30], [31].

Besides double-talk, another challenging issue to deal with in AEC is the loudspeaker nonlinearity, which makes the linear mixing model no longer appropriate [3], [5], i.e., the input of the linear mixing system is a nonlinear transform of the far-end signal. To deal with this problem, Volterra filters are often used [58], [59], [60]; but they require large memory and are computationally expensive and difficult to implement in most low-cost devices. Another simple yet efficient approach to model loudspeaker nonlinearities is through adding the so-called odd power series expansion [61], [62], [63] into the CTF framework, based on which several SBSS-based nonlinear AEC (SBSS-NAEC) algorithms have been proposed [57], [64]. These methods merge the CTF filter and the series expansion coefficients into a long vector, which represents the parameters to be identified [57], [64]. We will refer to this filter as the merged near-end signal extraction (MNE) filter in the rest of this work. The IP and EISS algorithms have been developed for estimating MNE filters, which have demonstrated promising NAEC performance [57], [64]. However, the computational complexity of such algorithms is quite high as the length of the MNE filter is generally large, which is proportional to the product of the CTF filter length and the number of series expansion coefficients.

To improve the efficiency and effectiveness of the IP and EISS algorithms, we propose in this work to represent the nonlinear echo in a bilinear form as discussed in previous studies [65], [66], which decouples the CTF filter and the series expansion coefficients. To the best of our knowledge, this is the first time

TABLE I
NOTATION OF IMPORTANT ACRONYMS

NAEC	Nonlinear acoustic echo cancellation
AIR	Acoustic impulse response
SBSS	Semi-blind source separation
MTF	Multiplicative transfer function
CTF	Convolutive transfer function
MNE	Merged near-end extraction
HEAD	Hybrid exact-approximate diagonalization
MM	Majorize-minimization
LCQP	Linear constrained quadratic programming
IP	Iterative projection
AIP	Alternating iterative projection
EISS	Element-wise iterative source steering
AEISS	Alternating element-wise iterative source steering

that such a form is explored in the frequency-domain SBSS-NAEC, though a similar model was investigated for time-domain adaptive filtering [67]. We then present a bilinear alternating optimization framework for estimating the model parameters. Based on this framework and also following the principles in [56], [57], [64], we derive an alternating iterative projection (AIP) algorithm and an alternating element-wise iterative source steering (AEISS) algorithm. Since the number of free parameters to estimate is less than those in the conventional IP and EISS methods, the derived two algorithms cannot only improve the AEC performance but also reduce the computational complexity.

The contributions of this work are as follows. Initially, we utilize a bilinear form to decouple the CTF filter and the nonlinear expansion coefficients, reformulating their estimation as two separate SBSS problems. Given the interdependence between the CTF filter and the nonlinear coefficients, we introduce a bilinear alternating optimization approach. This framework serves as the foundation for developing two efficient algorithms: AIP and AEISS. Simulations and experiments are carried out to validate the superior performance of AIP and AEISS. Important acronyms are detailed in Table I for future reference.

II. Signal Model

Consider a full-duplex communication scenario. The far-end signal is played back through a loudspeaker with some unknown nonlinearity. The loudspeaker signal is then convolved with the AIR as it travels from the loudspeaker to the near-end microphone, forming the so-called echo. Finally, the echo signal, along with the near-end signal, is captured by the near-end microphone and transmitted back to the far end. Mathematically, this process can be expressed at time t as

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

where $y(t)$ is the observed microphone signal, $s(t)$ is the near-end signal, $v(t) = a(t) \star f[x(t)]$ is the nonlinear acoustic echo, $a(t)$ stands for the AIR, $\star$ denotes the linear convolution, and

$f[\cdot]$ represents the loudspeaker responses, including both linear and nonlinear effects.

A. Odd Power Series Expansion

Instead of directly estimating the nonlinear system, a technique involving several linear systems is employed to approximate the nonlinearity. Subsequently, the problem is transformed into one of linear system identification, as discussed in previous studies [62], [63]. The Maclaurin series expansion emerges as a potent mathematical tool for approximating nonlinear functions. Given that memoryless nonlinearities prevalent in AEC applications are typically odd symmetric, the Maclaurin series expansion simplifies into the odd power series expansion. This approach has proven effective in modeling memoryless loudspeaker nonlinearities, as demonstrated in prior research [64], [67]. According to [56], [57], [61], [62], [63], [64], a typical memoryless nonlinearity can be expanded as

$$f\left[x(t)\right] = \sum_{n=1}^{N} b_n \, x^{2n-1}(t), \tag{2}$$

where N is the expansion order and b_n is the nth order odd power series expansion coefficient. It is noteworthy that Maclaurin series expansion can also handle nonsymmetric nonlinearities by doubling parameters. In such scenarios, the enhanced efficacy of the proposed algorithms becomes even more pronounced.

B. Convolutive Transfer Function Model

To accelerate SBSS-NAEC algorithms, the MTF model representing time-domain convolution with a frequency-domain multiplication is often adopted. However, the MTF model requires an STFT analysis window to be at least as long as the effective part of the linear echo path, which will inevitably introduce significant algorithmic delay, especially in highly reverberant environments [39]. In order to address this limitation, CTF models which represent the time-domain convolution with a frequency-domain convolution is adopted [40], [41], [42]. CTF models are not restricted by the length of STFT window and, therefore, can achieve a compromise between the computational complexity and algorithmic delay [40]. With this model [41], the microphone signal can be expressed as [57], [64]

$$Y_{i,j} = S_{i,j} + \underbrace{\sum_{l=1}^{L}\sum_{n=1}^{N} b_n \, A_{i,j,l} X_{n,i,j-l+1}}_{V_{i,j}}, \tag{3}$$

where $i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, J\}$ are the frequency and time-frame indices, respectively, L is the length of the CTF subband filter, $Y_{i,j}$, $S_{i,j}$, $V_{i,j}$, and $X_{n,i,j}$ are the STFT-domain representations of $y(t)$, $s(t)$, $v(t)$, and $x^{2n-1}(t)$, respectively, and $A_{i,j,l}$ is a CTF filter coefficient representing the linear echo path in the CTF domain. Generally, for a given reverberation level, if shorter algorithmic delay is desired, a shorter STFT window has to be used. Consequently, a larger CTF filter length L has to be chosen [40], [42], which however inevitably leads to increase of the computational complexity.

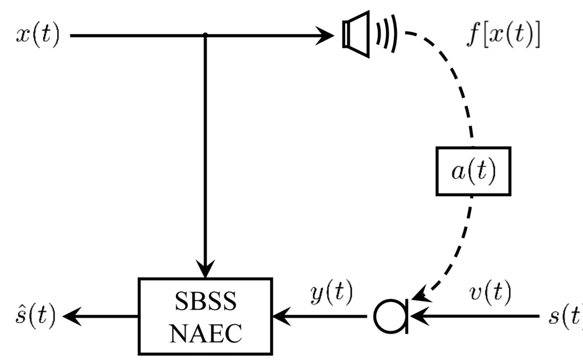


Fig. 1. Illustration of NAEC in full duplex communications.

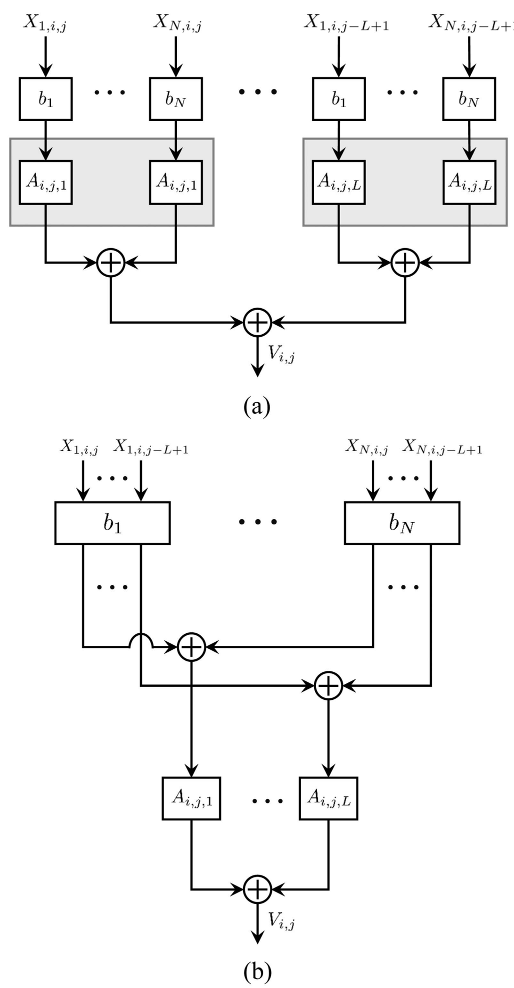


Fig. 2. Illustration of the nonlinear echo model: (a) merged nonlinear echo model and (b) bilinear nonlinear echo model.

C. Merged Nonlinear Echo Model

In the conventional SBSS-NAEC model [57], [64], the nonlinear coefficients are merged into the linear echo path as illustrated in Fig. 2(a). Mathematically, this process can be expressed as

$$V_{i,j} = \sum_{l=1}^{L}\sum_{n=1}^{N} b_n \, A_{i,j,l} X_{n,i,j-l+1}$$

$$
\begin{aligned}
&= \sum_{l=1}^{L} \sum_{n=1}^{N} A^{\mathrm{m}}_{n,i,j,l} X_{n,i,j-l+1} \\
&= \sum_{l=1}^{L} \left(\mathbf{a}^{\mathrm{m}}_{i,j,l}\right)^T \mathbf{x}_{i,j-l+1} \\
&= \left(\tilde{\mathbf{a}}^{\mathrm{m}}_{i,j}\right)^T \tilde{\mathbf{x}}_{i,j}, \qquad (4)
\end{aligned}
$$

where $A^{\mathrm{m}}_{n,i,j,l} = b_n A_{i,j,l}$, the superscripts m and T denote, respectively, the merged model and transpose operator, and

$$\mathbf{a}^{\mathrm{m}}_{i,j,l} = \left[A^{\mathrm{m}}_{1,i,j,l} \quad A^{\mathrm{m}}_{2,i,j,l} \quad \ldots \quad A^{\mathrm{m}}_{N,i,j,l}\right]^T, \tag{5}$$

$$\tilde{\mathbf{a}}^{\mathrm{m}}_{i,j} = \left[\left(\mathbf{a}^{\mathrm{m}}_{i,j,1}\right)^T \quad \left(\mathbf{a}^{\mathrm{m}}_{i,j,2}\right)^T \quad \ldots \quad \left(\mathbf{a}^{\mathrm{m}}_{i,j,L}\right)^T\right]^T, \tag{6}$$

$$\mathbf{x}_{i,j-l} = \left[X_{1,i,j-l} \quad X_{2,i,j-l} \quad \ldots \quad X_{N,i,j-l}\right]^T, \tag{7}$$

$$\tilde{\mathbf{x}}_{i,j} = \left[\mathbf{x}^T_{i,j} \quad \mathbf{x}^T_{i,j-1} \quad \ldots \quad \mathbf{x}^T_{i,j-L+1}\right]^T. \tag{8}$$

As seen, in the merged model, the nonlinear echo is represented with $N \times I \times L$ parameters.

D. Bilinear Nonlinear Echo Model

Alternatively, as can be seen from (3), the nonlinear echo can be reformulated as a bilinear form of the odd power series expansion coefficients vector $\mathbf{b}$ and the CTF filter vector $\mathbf{a}_{i,j}$:

$$
\begin{aligned}
V_{i,j} &= \mathbf{a}^T_{i,j}\, \mathbf{X}_{i,j}\, \mathbf{b}, \\
&= \mathbf{a}^T_{i,j} \mathbf{x}^{\mathrm{a}}_{i,j}, \qquad (9)
\end{aligned}
$$

or, equivalently,

$$
\begin{aligned}
V_{i,j} &= \mathbf{b}^T\, \mathbf{X}^T_{i,j}\, \mathbf{a}_{i,j} \\
&= \mathbf{b}^T \mathbf{x}^{\mathrm{b}}_{i,j}, \qquad (10)
\end{aligned}
$$

where

$$\mathbf{a}_{i,j} = \left[A_{i,j,1} \quad A_{i,j,2} \quad \ldots \quad A_{i,j,L}\right]^T, \tag{11}$$

$$\mathbf{b} = \left[b_1 \quad b_2 \quad \ldots \quad b_N\right]^T, \tag{12}$$

and

$$\mathbf{X}_{i,j} = \begin{bmatrix} X_{1,i,j} & X_{2,i,j} & \cdots & X_{N,i,j} \\ X_{1,i,j-1} & X_{2,i,j-1} & \cdots & X_{N,i,j-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1,i,j-L+1} & X_{2,i,j-L+1} & \cdots & X_{N,i,j-L+1} \end{bmatrix}. \tag{13}$$

The two transformed signal vectors $\mathbf{x}^{\mathrm{a}}_{i,j}$ and $\mathbf{x}^{\mathrm{b}}_{i,j}$ are of the following form:

$$\mathbf{x}^{\mathrm{a}}_{i,j} = \mathbf{X}_{i,j} \mathbf{b}, \tag{14}$$

$$\mathbf{x}^{\mathrm{b}}_{i,j} = \mathbf{X}^T_{i,j} \mathbf{a}_{i,j}. \tag{15}$$

By decoupling the series expansion parameters $\mathbf{b}$ and the CTF filter coefficients $\mathbf{a}_{i,j}$ from each other by the bilinear nonlinear echo model, the number of free parameters for every time frame is reduced to $N + I \times L$. This bilinear nonlinear echo model is illustrated in Fig. 2(b).

E. Probabilistic Signal Model

In the rest of this paper, we model the near-end signal with a generalized Gaussian distribution, which has been widely used in the literature [47], [48], [56], [57], [64]:

$$p\left(\mathbf{s}_j\right) \propto \exp\left[-\left(\frac{\|\mathbf{s}_j\|_2}{\gamma}\right)^{\beta}\right], \tag{16}$$

where

$$\mathbf{s}_j = \left[S_{1,j} \quad S_{2,j} \quad \ldots \quad S_{I,j}\right]^T \tag{17}$$

and $||\cdot||_2$ denotes the ℓ_2 norm. We assume that $\gamma > 0$ and $0 < \beta \leq 2$ to obtain supergaussian distributions in order to use the well-known majorize-minimization (MM) method [68].

III. Conventional Algorithms

In this section, we give a brief overview of the conventional IP and EISS algorithms, which are based on the merged nonlinear echo model. To formulate the NAEC problem as one of SBSS, the following mixing model is considered

$$\tilde{\mathbf{y}}^{\mathrm{m}}_{i,j} = \mathbf{H}^{\mathrm{m}}_{i,j} \tilde{\mathbf{s}}^{\mathrm{m}}_{i,j}, \tag{18}$$

where

$$\tilde{\mathbf{y}}^{\mathrm{m}}_{i,j} = \left[Y_{i,j} \quad \tilde{\mathbf{x}}^T_{i,j}\right]^T, \tag{19}$$

$$\tilde{\mathbf{s}}^{\mathrm{m}}_{i,j} = \left[S_{i,j} \quad \tilde{\mathbf{x}}^T_{i,j}\right]^T \tag{20}$$

are two signal vectors, and

$$\mathbf{H}^{\mathrm{m}}_{i,j} = \begin{bmatrix} 1 & \left(\tilde{\mathbf{a}}^{\mathrm{m}}_{i,j}\right)^T \\ \mathbf{0}_{NL\times 1} & \mathbf{I}_{NL} \end{bmatrix} \tag{21}$$

is the mixing matrix, with $\mathbf{0}_{NL\times 1}$ being an all-zero column vector of length NL and $\mathbf{I}_{NL}$ being the identity matrix of size $NL \times NL$. Now the problem can be solved from a signal separation perspective. The near-end signal can then be recovered through the following inverse system:

$$\mathbf{W}^{\mathrm{m}}_{i,j} = \begin{bmatrix} 1 & -\left(\hat{\tilde{\mathbf{a}}}^{\mathrm{m}}_{i,j}\right)^T \\ \mathbf{0}_{NL\times 1} & \mathbf{I}_{NL} \end{bmatrix}, \tag{22}$$

where $\hat{\tilde{\mathbf{a}}}^{\mathrm{m}}_{i,j}$ is the estimate of $\tilde{\mathbf{a}}^{\mathrm{m}}_{i,j}$. The MNE filter defined as

$$\mathbf{w}^{\mathrm{m}}_{i,j} = \left[1 \quad -\left(\hat{\tilde{\mathbf{a}}}^{\mathrm{m}}_{i,j}\right)^T\right]^H, \tag{23}$$

where $(\cdot)^H$ denotes conjugate transpose, is used to extract the near-end signal, i.e.,

$$\hat{S}_{i,j} = \left(\mathbf{w}^{\mathrm{m}}_{i,j}\right)^H \tilde{\mathbf{y}}^{\mathrm{m}}_{i,j}. \tag{24}$$

It is important to emphasize that this signal separation issue qualifies as an SBSS problem as the far-end signal is precisely known. Now exploiting the mutual independence between the near-end and far-end signals, one can derive the following recursive negative log-likelihood function [57]:

$$\mathcal{L}^{\mathrm{m}}_{i,j} = -\frac{1}{\sum_{j'=1}^{j} (\eta^{\mathrm{m}})^{j-j'}} \sum_{j'=1}^{j} (\eta^{\mathrm{m}})^{j-j'} \log p\left(\mathbf{s}_{j'}\right)$$

$$-2\sum_{i=1}^{I}\log\left|\det \mathbf{W}_{i,j}^{\mathrm{m}}\right|, \tag{25}$$

where $\eta^{\mathrm{m}} \in (0,1)$ is a forgetting factor. With the MM method [47], [48], the following auxiliary function can be constructed

$$\mathcal{L}_{i,j}^{\mathrm{m},+} = \sum_{i=1}^{I}\left(\mathbf{w}_{i,j}^{\mathrm{m}}\right)^{H}\mathbf{G}_{i,j}^{\mathrm{m}}\mathbf{w}_{i,j}^{\mathrm{m}} - 2\sum_{i=1}^{I}\log\left|\det \mathbf{W}_{i,j}^{\mathrm{m}}\right|, \tag{26}$$

where

$$\mathbf{G}_{i,j}^{\mathrm{m}} = \eta^{\mathrm{m}}\mathbf{G}_{i,j-1}^{\mathrm{m}} + (1-\eta^{\mathrm{m}})\,\varphi\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right)\tilde{\mathbf{y}}_{i,j}^{\mathrm{m}}(\tilde{\mathbf{y}}_{i,j}^{\mathrm{m}})^{H} \tag{27}$$

is the auxiliary matrix and

$$\varphi\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right) = (\sigma_{\mathrm{s},j}^{\mathrm{m}})^{\beta-2}, \tag{28}$$

$$\sigma_{\mathrm{s},j}^{\mathrm{m}} = \sqrt{\sum_{i=1}^{I}\left|(\mathbf{w}_{i,j-1}^{\mathrm{m}})^{H}\tilde{\mathbf{y}}_{i,j}^{\mathrm{m}}\right|^{2}}, \tag{29}$$

is a weighting function. It can be shown that [47], [48]

$$\mathcal{L}_{i,j}^{\mathrm{m},+} \geq \mathcal{L}_{i,j}^{\mathrm{m}}, \tag{30}$$

with equality if and only if $\mathbf{w}_{i,j} = \mathbf{w}_{i,j-1}$. The derivation of the auxiliary function is presented in Appendix A. Due to the structure of $\mathbf{W}_{i,j}^{\mathrm{m}}$ given in (22), (26) is a simplified HEAD problem [46], which has been well studied in the literature of BSS [47], [51], [52]. Accordingly, the MNE filter (23) can be optimized adaptively by constructing and minimizing (26). Then the near-end signal is extracted according to (24).

A. Iterative Projection

By equating the Wirtinger derivative of (26) with respect to $(\mathbf{w}_{i,j}^{\mathrm{m}})^{*}$ to 0, where $(\cdot)^{*}$ denotes complex conjugation, one obtains

$$\mathbf{G}_{i,j}^{\mathrm{m}}\mathbf{w}_{i,j}^{\mathrm{m}} = \left(\mathbf{W}_{i,j}^{\mathrm{m}}\right)^{-1}\mathbf{e}_{NL+1}, \tag{31}$$

where $\mathbf{e}_{NL+1}$ is a unitary vector of length $NL+1$, with the first element being 1. Then the IP-based update rule is obtained as [47], [48]

$$\mathbf{w}_{i,j}^{\mathrm{m}} = \left(\mathbf{W}_{i,j}^{\mathrm{m}}\mathbf{G}_{i,j}^{\mathrm{m}}\right)^{-1}\mathbf{e}_{NL+1}. \tag{32}$$

Taking the structure of $\mathbf{W}_{i,j}^{\mathrm{m}}$ into account, the update rule can be simplified as [56], [64]

$$\mathbf{w}_{i,j}^{\mathrm{m}} = \left(\mathbf{G}_{i,j}^{\mathrm{m}}\right)^{-1}\mathbf{e}_{NL+1}. \tag{33}$$

Now, we ensure the structure of the MNE filter (23), to be specific, the first element should be 1, by

$$\mathbf{w}_{i,j}^{\mathrm{m}} := \frac{\mathbf{w}_{i,j}^{\mathrm{m}}}{W_{i,j,1}^{\mathrm{m}}}, \tag{34}$$

where $W_{i,j,1}^{\mathrm{m}}$ is the first element of $\mathbf{w}_{i,j}^{\mathrm{m}}$ and $:=$ denotes assignment.

One may notice that the IP algorithm updates the MNE filter in a two-step manner, i.e., update and normalization, which does not fully take advantage of the structure of the unit upper triangular structure of the demixing matrix (22).

B. Element-Wise Iterative Source Steering

A more efficient update strategy is given by the EISS algorithm [57], which is modified from the rank-1 update rule proposed in [51]. For EISS, the MNE filter is updated element-wisely as

$$W_{i,j,k}^{\mathrm{m}} = \begin{cases} W_{i,j-1,k}^{\mathrm{m}} - U_{i,j,k}^{\mathrm{m}}, & k=1 \\ \left(1-U_{i,j,1}^{\mathrm{m}}\right)W_{i,j-1,k}^{\mathrm{m}} - U_{i,j,k}^{\mathrm{m}}, & \text{else} \end{cases}, \tag{35}$$

where $W_{i,j,k}^{\mathrm{m}}$ is the k-th element of the MNE filter $\mathbf{w}_{i,j}^{\mathrm{m}}$ and $U_{i,j,k}^{\mathrm{m}}$ is a steering stepsize yet to be determined. Substituting (35) into (26) gives the following auxiliary function:

$$\mathcal{L}_{i,j}^{\mathrm{m},+} = -2\log\left|1-(U_{i,j,1}^{\mathrm{m}})^{*}\right| + \left(\mathbf{w}_{i,j}^{\mathrm{m}} - \tilde{\mathbf{u}}_{i,j}^{\mathrm{m}}\right)^{H}\mathbf{G}_{j}^{\mathrm{m}}\left(\mathbf{w}_{i,j}^{\mathrm{m}} - \tilde{\mathbf{u}}_{i,j}^{\mathrm{m}}\right), \tag{36}$$

where

$$\tilde{\mathbf{u}}_{i,j}^{\mathrm{m}} = \left[U_{i,j,1}^{\mathrm{m}} \quad U_{i,j,1}^{\mathrm{m}}W_{i,j-1,2}^{\mathrm{m}} + U_{i,j,2}^{\mathrm{m}} \quad \ldots \quad U_{i,j,1}^{\mathrm{m}}W_{i,j-1,NL+1}^{\mathrm{m}} + U_{i,j,NL+1}^{\mathrm{m}}\right]^{T} \tag{37}$$

is a vector containing the parameters to be estimated. Then, forcing the derivative with respect to $(U_{i,j,k}^{\mathrm{m}})^{*}$ to 0 gives

$$U_{i,j,k}^{\mathrm{m}} = \begin{cases} 1-\left[\left(\mathbf{w}_{i,j}^{\mathrm{m}}\right)^{H}\mathbf{G}_{i,j}^{\mathrm{m}}\mathbf{w}_{i,j}^{\mathrm{m}}\right]^{-\frac{1}{2}}, & k=1 \\ \dfrac{\left(\mathbf{g}_{i,j,k}^{\mathrm{m}}\right)^{H}\mathbf{w}_{i,j}^{\mathrm{m}}}{G_{i,j,k,k}^{\mathrm{m}}}, & \text{else} \end{cases}, \tag{38}$$

where $\mathbf{g}_{i,j,k}^{\mathrm{m}}$ is the kth column of matrix $\mathbf{G}_{i,j}^{\mathrm{m}}$ and $G_{i,j,k,k}^{\mathrm{m}}$ is the kth diagonal element of $\mathbf{G}_{i,j}^{\mathrm{m}}$. After updating the MNE filter with (38) and (35), the normalization given in (34) is performed.

Again, the EISS algorithm updates the MNE filter coefficients in the two-step manner, which does not take the unit upper triangular structure of the demixing matrix (22) into full consideration.

IV. Proposed Methods

As can be seen, the MNE filter length is $NL+1$. Consequently, the computational complexity for identifying the merged nonlinear echo model is high. To address this issue, we propose two improved algorithms based on the previously discussed bilinear nonlinear echo model, which take advantage of the structure of the demixing system to update the filter coefficients with a one-step strategy. Note that the two-step update strategy in the IP and EISS algorithms can also be adopted to the presented methods; but the dimension of the related matrices is higher, which makes the update process computationally more expensive.

Firstly, if the series expansion coefficients $\mathbf{b}_j$ are fixed, we can formulate the identification of the CTF filter $\mathbf{a}_{i,j}$ as

$$\tilde{\mathbf{y}}_{i,j}^{\mathrm{a}} = \mathbf{H}_{i,j}^{\mathrm{a}}\tilde{\mathbf{s}}_{i,j}^{\mathrm{a}}, \tag{39}$$

where

$$\tilde{\mathbf{y}}_{i,j}^{\mathrm{a}} = \left[Y_{i,j} \quad (\mathbf{x}_{i,j}^{\mathrm{a}})^{T}\right]^{T}, \tag{40}$$

$$\tilde{\mathbf{s}}_{i,j}^{\mathrm{a}} = \left[S_{i,j} \quad (\mathbf{x}_{i,j}^{\mathrm{a}})^{T}\right]^{T} \tag{41}$$

are two signal vectors, and

$$\mathbf{H}^{\mathrm{a}}_{i,j} = \begin{bmatrix} 1 & \mathbf{a}^T_{i,j} \\ \mathbf{0}_{L\times 1} & \mathbf{I}_L \end{bmatrix} \quad (42)$$

is the mixing matrix. The near-end signal is extracted by applying the demixing matrix:

$$\mathbf{W}^{\mathrm{a}}_{i,j} = \begin{bmatrix} 1 & -\hat{\mathbf{a}}^T_{i,j} \\ \mathbf{0}_{L\times 1} & \mathbf{I}_L \end{bmatrix}, \quad (43)$$

where $\hat{\mathbf{a}}^T_{i,j}$ is the estimate of $\mathbf{a}^T_{i,j}$. With the filter

$$\mathbf{w}^{\mathrm{a}}_{i,j} = \begin{bmatrix} 1 & -\hat{\mathbf{a}}^T_{i,j} \end{bmatrix}^H, \quad (44)$$

the near-end signal can be extracted

$$\hat{S}_{i,j} = \left(\mathbf{w}^{\mathrm{a}}_{i,j}\right)^H \tilde{\mathbf{y}}^{\mathrm{a}}_{i,j}. \quad (45)$$

To estimate $\mathbf{a}_{i,j}$, the following cost function is considered

$$\mathcal{L}^{\mathrm{a}}_{i,j} = -\frac{1}{\sum_{j'=1}^{j}(\eta^{\mathrm{a}})^{j-j'}} \sum_{j'=1}^{j} (\eta^{\mathrm{a}})^{j-j'} \log p\left(\mathbf{s}_{j'}\right) - 2\sum_{i=1}^{I} \log\left|\det \mathbf{W}^{\mathrm{a}}_{i,j}\right|, \quad (46)$$

where $\eta^{\mathrm{a}} \in (0,1)$ is a forgetting factor. As the first element of $\mathbf{w}^{\mathrm{a}}_{i,j}$ is fixed to 1, we obtain due to the one-step strategy:

$$\det \mathbf{W}^{\mathrm{a}}_{i,j} = 1. \quad (47)$$

Additionally, with the majorize-minimization (MM) method, the following auxiliary function can be constructed

$$\mathcal{L}^{\mathrm{a},+}_{i,j} = \sum_{i=1}^{I} \left(\mathbf{w}^{\mathrm{a}}_{i,j}\right)^H \mathbf{G}^{\mathrm{a}}_{i,j} \mathbf{w}^{\mathrm{a}}_{i,j}, \quad (48)$$

where the auxiliary matrix $\mathbf{G}^{\mathrm{a}}_{i,j}$ can be partitioned as

$$\mathbf{G}^{\mathrm{a}}_{i,j} = \begin{bmatrix} \left(\sigma^{\mathrm{a}}_{\mathrm{y},i,j}\right)^2 & (\mathbf{q}^{\mathrm{a}}_{i,j})^H \\ \mathbf{q}^{\mathrm{a}}_{i,j} & \mathbf{R}^{\mathrm{a}}_{i,j} \end{bmatrix}, \quad (49)$$

where

$$\left(\sigma^{\mathrm{a}}_{\mathrm{y},i,j}\right)^2 = \eta^{\mathrm{a}} \left(\sigma^{\mathrm{a}}_{\mathrm{y},i,j-1}\right)^2 + (1-\eta^{\mathrm{a}})\varphi\left(\sigma^{\mathrm{a}}_{\mathrm{s},j}\right)|Y_{i,j}|^2, \quad (50)$$

$$\mathbf{q}^{\mathrm{a}}_{i,j} = \eta^{\mathrm{a}}\mathbf{q}^{\mathrm{a}}_{i,j-1} + (1-\eta^{\mathrm{a}})\varphi\left(\sigma^{\mathrm{a}}_{\mathrm{s},j}\right) Y^*_{i,j}\mathbf{x}^{\mathrm{a}}_{i,j}, \quad (51)$$

$$\mathbf{R}^{\mathrm{a}}_{i,j} = \eta^{\mathrm{a}}\mathbf{R}^{\mathrm{a}}_{i,j-1} + (1-\eta^{\mathrm{a}})\varphi\left(\sigma^{\mathrm{a}}_{\mathrm{s},j}\right) \mathbf{x}^{\mathrm{a}}_{i,j}(\mathbf{x}^{\mathrm{a}}_{i,j})^H, \quad (52)$$

with

$$\varphi\left(\sigma^{\mathrm{a}}_{\mathrm{s},j}\right) = (\sigma^{\mathrm{a}}_{\mathrm{s},j})^{\beta-2}, \quad (53)$$

and

$$\sigma^{\mathrm{a}}_{\mathrm{s},j} = \sqrt{\sum_{i=1}^{I} \left|(\mathbf{w}^{\mathrm{a}}_{i,j-1})^H \tilde{\mathbf{y}}^{\mathrm{a}}_{i,j}\right|^2}. \quad (54)$$

Analogously to (30), one can verify that

$$\mathcal{L}^{\mathrm{a},+}_{i,j} \geq \mathcal{L}^{\mathrm{a}}_{i,j}. \quad (55)$$

Now, $\hat{\mathbf{a}}_{i,j}$ can be optimized by solving

$$\hat{\mathbf{a}}_{i,j} = \operatorname{argmin} \mathcal{L}^{\mathrm{a},+}_{i,j}, \quad \text{s.t. } \left(\mathbf{w}^{\mathrm{a}}_{i,j}\right)^H \mathbf{e}_{L+1} = 1. \quad (56)$$

After $\hat{\mathbf{a}}_{i,j}$ are updated, one can construct $\mathbf{x}^{\mathrm{b}}_{i,j}$ according to (15). The series expansion coefficients can then be identified as

$$\tilde{\mathbf{y}}^{\mathrm{b}}_{i,j} = \mathbf{H}^{\mathrm{b}} \tilde{\mathbf{s}}^{\mathrm{b}}_{i,j}, \quad (57)$$

where

$$\tilde{\mathbf{y}}^{\mathrm{b}}_{i,j} = \begin{bmatrix} Y_{i,j} & \left(\mathbf{x}^{\mathrm{b}}_{i,j}\right)^T \end{bmatrix}^T, \quad (58)$$

$$\tilde{\mathbf{s}}^{\mathrm{b}}_{i,j} = \begin{bmatrix} S_{i,j} & \left(\mathbf{x}^{\mathrm{b}}_{i,j}\right)^T \end{bmatrix}^T, \quad (59)$$

$$\mathbf{H}^{\mathrm{b}} = \begin{bmatrix} 1 & \mathbf{b}^T \\ \mathbf{0}_{N\times 1} & \mathbf{I}_N \end{bmatrix}. \quad (60)$$

The estimated demixing matrix is

$$\mathbf{W}^{\mathrm{b}}_{j} = \begin{bmatrix} 1 & -\hat{\mathbf{b}}^T_j \\ \mathbf{0}_{N\times 1} & \mathbf{I}_N \end{bmatrix}, \quad (61)$$

where $\hat{\mathbf{b}}_j$ is the estimate of $\mathbf{b}$ at time-frame j. Using the following filter:

$$\mathbf{w}^{\mathrm{b}}_{j} = \begin{bmatrix} 1 & -\hat{\mathbf{b}}^T_j \end{bmatrix}^H, \quad (62)$$

one can extract the near-end signal as

$$\hat{S}_{i,j} = \left(\mathbf{w}^{\mathrm{b}}_{j}\right)^H \tilde{\mathbf{y}}^{\mathrm{b}}_{i,j}. \quad (63)$$

Now, the recursive negative log-likelihood function for $\hat{\mathbf{b}}_j$ is written as

$$\mathcal{L}^{\mathrm{b}}_{j} = -\frac{1}{\sum_{j'=1}^{j}(\eta^{\mathrm{b}})^{j-j'}} \sum_{j'=1}^{j} (\eta^{\mathrm{b}})^{j-j'} \log p\left(\mathbf{s}_{j'}\right) - 2I \log\left|\det \mathbf{W}^{\mathrm{b}}_{j}\right|, \quad (64)$$

where $\eta^{\mathrm{b}} \in (0,1)$ is a forgetting factor for estimating the series expansion coefficients. Analogously, with the one-step strategy, we have

$$\det \mathbf{W}^{\mathrm{b}}_{j} = 1. \quad (65)$$

Therefore, with the MM method, the following auxiliary function is constructed

$$\mathcal{L}^{\mathrm{b},+}_{j} = \left(\mathbf{w}^{\mathrm{b}}_{j}\right)^H \bar{\mathbf{G}}^{\mathrm{b}}_{j} \mathbf{w}^{\mathrm{b}}_{j}. \quad (66)$$

The auxiliary matrix $\bar{\mathbf{G}}^{\mathrm{b}}_{j}$ can again be partitioned as

$$\bar{\mathbf{G}}^{\mathrm{b}}_{j} = \begin{bmatrix} \left(\bar{\sigma}^{\mathrm{b}}_{\mathrm{y},j}\right)^2 & (\bar{\mathbf{q}}^{\mathrm{b}}_{j})^H \\ \bar{\mathbf{q}}^{\mathrm{b}}_{j} & \bar{\mathbf{R}}^{\mathrm{b}}_{j} \end{bmatrix}, \quad (67)$$

where

$$\left(\bar{\sigma}^{\mathrm{b}}_{\mathrm{y},j}\right)^2 = \eta^{\mathrm{b}} \left(\bar{\sigma}^{\mathrm{b}}_{\mathrm{y},j-1}\right)^2 + \frac{\left(1-\eta^{\mathrm{b}}\right)\varphi\left(\sigma^{\mathrm{b}}_{\mathrm{s},j}\right)}{I} \sum_{i=1}^{I} |Y_{i,j}|^2, \quad (68)$$

$$\bar{\mathbf{q}}^{\mathrm{b}}_{j} = \eta^{\mathrm{b}}\bar{\mathbf{q}}^{\mathrm{b}}_{j-1} + \frac{\left(1-\eta^{\mathrm{b}}\right)\varphi\left(\sigma^{\mathrm{b}}_{\mathrm{s},j}\right)}{I} \sum_{i=1}^{I} Y^*_{i,j}\mathbf{x}^{\mathrm{b}}_{i,j}, \quad (69)$$

$$\bar{\mathbf{R}}_j^{\mathrm{b}} = \eta^{\mathrm{b}} \bar{\mathbf{R}}_{j-1}^{\mathrm{b}} + \frac{\left(1-\eta^{\mathrm{b}}\right) \varphi\left(\sigma_{\mathrm{s},j}^{\mathrm{b}}\right)}{I} \sum_{i=1}^{I} \mathbf{x}_{i,j}^{\mathrm{b}} \left(\mathbf{x}_{i,j}^{\mathrm{b}}\right)^H, \tag{70}$$

with

$$\varphi\left(\sigma_{\mathrm{s},j}^{\mathrm{b}}\right) = \left(\sigma_{\mathrm{s},j}^{\mathrm{b}}\right)^{\beta-2}, \tag{71}$$

and

$$\sigma_{\mathrm{s},j}^{\mathrm{b}} = \sqrt{\sum_{i=1}^{I} \left| \left(\mathbf{w}_{j-1}^{\mathrm{b}}\right)^H \tilde{\mathbf{y}}_{i,j}^{\mathrm{b}} \right|^2}. \tag{72}$$

Then the nonlinear coefficients can be estimated by solving

$$\hat{\mathbf{b}}_j = \operatorname{argmin} \mathcal{L}_j^{\mathrm{b},+}, \quad \text{s.t. } \left(\mathbf{w}_j^{\mathrm{b}}\right)^H \mathbf{e}_{N+1} = 1. \tag{73}$$

Now, the entire system can be identified by iteratively and alternatingly updating $\hat{\mathbf{w}}_{i,j}^{\mathrm{a}}$ and $\hat{\mathbf{w}}_j^{\mathrm{b}}$. Finally, the near-end signal can be extracted by applying (63).

A. Alternating Iterative Projection

We now adopt the IP algorithm for our bilinear alternating optimization scheme by directly solving (56) and (73). Note that as $\det(\mathbf{W}_{i,j}^{\mathrm{a}}) = \det(\mathbf{W}_j^{\mathrm{b}}) = 1$, the HEAD problem degenerates into one of linear constrained quadratic programming (LCQP) [69], [70], [71], which has been widely studied in minimum variance distortionless response (MVDR) filter [69], [72], [73] and linearly constrained minimum variance (LCMV) filter [69], [74], [75], [76] optimization.

Firstly, we use the estimate $\hat{\mathbf{b}}_j$ from the previous frame to construct $\mathbf{x}_{i,j}^{\mathrm{a}}$ and update the associated statistics. The solution of (56) is given as [72]

$$\mathbf{w}_{i,j}^{\mathrm{a}} = \frac{\left(\mathbf{G}_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{e}_{L+1}}{\mathbf{e}_{L+1}^T \left(\mathbf{G}_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{e}_{L+1}}. \tag{74}$$

Assuming that both $\mathbf{G}_{i,j}^{\mathrm{a}}$ and $\mathbf{R}_{i,j}^{\mathrm{a}}$ are invertible, one can express the inverse of $\mathbf{G}_{i,j}^{\mathrm{a}}$ as

$$\left(\mathbf{G}_{i,j}^{\mathrm{a}}\right)^{-1} = \begin{bmatrix} \left(S_{i,j}^{\mathrm{a}}\right)^{-1} & -\left(S_{i,j}^{\mathrm{a}}\right)^{-1} \left(\mathbf{c}_{i,j}^{\mathrm{a}}\right)^H \\ -\left(S_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{c}_{i,j}^{\mathrm{a}} & \left(\mathbf{R}_{i,j}^{\mathrm{a}}\right)^{-1} + \left(S_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{D}_{i,j}^{\mathrm{a}} \end{bmatrix}, \tag{75}$$

where

$$S_{i,j}^{\mathrm{a}} = \left(\sigma_{\mathrm{y},i,j}^{\mathrm{a}}\right)^2 - \left(\mathbf{q}_{i,j}^{\mathrm{a}}\right)^H \left(\mathbf{R}_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{q}_{i,j}^{\mathrm{a}} \tag{76}$$

is the Schur complement of $\mathbf{R}_{i,j}^{\mathrm{a}}$ in $\mathbf{G}_{i,j}^{\mathrm{a}}$ [77], [78], [79] and

$$\mathbf{c}_{i,j}^{\mathrm{a}} = \left(\mathbf{R}_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{q}_{i,j}^{\mathrm{a}}, \tag{77}$$

$$\mathbf{D}_{i,j}^{\mathrm{a}} = \mathbf{c}_{i,j}^{\mathrm{a}} \left(\mathbf{c}_{i,j}^{\mathrm{a}}\right)^H. \tag{78}$$

Substituting (75) into (74), we have

$$\left(S_{i,j}^{\mathrm{a}}\right)^{-1} \begin{bmatrix} 1 \\ -\mathbf{a}_{i,j}^* \end{bmatrix} = \begin{bmatrix} \left(S_{i,j}^{\mathrm{a}}\right)^{-1} \\ -\left(S_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{c}_{i,j}^{\mathrm{a}} \end{bmatrix}. \tag{79}$$

Therefore, the CTF filter can be updated as

$$\hat{\mathbf{a}}_{i,j} = \left[\left(\mathbf{R}_{i,j}^{\mathrm{a}}\right)^{-1} \mathbf{q}_{i,j}^{\mathrm{a}}\right]^*. \tag{80}$$

Algorithm 1: Alternating iterative projection algorithm.

1: Setting forgetting factors η^{a} and η^{b}
 Initial parameters
 $\mathbf{X}_{i,0} = \mathbf{0}_{L\times N}$, $\hat{\mathbf{a}}_{i,0} = \mathbf{0}_{L\times 1}$, $\hat{\mathbf{b}}_0 = [1 \quad \mathbf{0}_{(N-1)\times 1}^T]^T$,
 $\mathbf{q}_{i,0}^{\mathrm{a}} = \mathbf{0}_{L\times 1}$, $\bar{\mathbf{q}}_0^{\mathrm{b}} = \mathbf{0}_{N\times 1}$,
 $\mathbf{R}_{i,0}^{\mathrm{a}} = 10^{-4}\mathbf{I}_L$, $\bar{\mathbf{R}}_0^{\mathrm{b}} = 10^{-4}\mathbf{I}_N$
2: **for** $j = 1$; $j < J$; $j = j+1$ **do**
3: Insert $X_{n,i,j}$ into $\mathbf{X}_{i,j-1}$ to get $\mathbf{X}_{i,j}$
4: Update CTF filter associated statistics with (14), (40), (51), (52), (53), (54)
 $\mathbf{x}_{i,j}^{\mathrm{a}} = \mathbf{X}_{i,j}\hat{\mathbf{b}}_{j-1}$,
 $\tilde{\mathbf{y}}_{i,j}^{\mathrm{a}} = [\ Y_{i,j} \quad (\mathbf{x}_{i,j}^{\mathrm{a}})^T\ ]^T$
 $\sigma_{\mathrm{s},j}^{\mathrm{a}} = \sqrt{\sum_{i=1}^{I} |(\mathbf{w}_{i,j-1}^{\mathrm{a}})^H \tilde{\mathbf{y}}_{i,j}^{\mathrm{a}}|^2}$
 $\varphi(\sigma_{\mathrm{s},j}^{\mathrm{a}}) = (\sigma_{\mathrm{s},j}^{\mathrm{a}})^{\beta-2}$
 $\mathbf{q}_{i,j}^{\mathrm{a}} = \eta^{\mathrm{a}}\mathbf{q}_{i,j-1}^{\mathrm{a}} + (1-\eta^{\mathrm{a}})\varphi(\sigma_{\mathrm{s},j}^{\mathrm{a}}) Y_{i,j}^* \mathbf{x}_{i,j}^{\mathrm{a}}$
 $\mathbf{R}_{i,j}^{\mathrm{a}} = \eta^{\mathrm{a}}\mathbf{R}_{i,j-1}^{\mathrm{a}} + (1-\eta^{\mathrm{a}})\varphi(\sigma_{\mathrm{s},j}^{\mathrm{a}}) \mathbf{x}_{i,j}^{\mathrm{a}} (\mathbf{x}_{i,j}^{\mathrm{a}})^H$
5: Update CTF filter with (80)
 $\hat{\mathbf{a}}_{i,j} = [(\mathbf{R}_{i,j}^{\mathrm{a}})^{-1}\mathbf{q}_{i,j}^{\mathrm{a}}]^*$
6: Update series expansion coefficients associated statistics with (15), (58), (69), (70), (71), (72)
 $\mathbf{x}_{i,j}^{\mathrm{b}} = \mathbf{X}_{i,j}^T \hat{\mathbf{a}}_{i,j}$
 $\tilde{\mathbf{y}}_{i,j}^{\mathrm{b}} = [Y_{i,j} \quad (\mathbf{x}_{i,j}^{\mathrm{b}})^T]^T$
 $\sigma_{\mathrm{s},j}^{\mathrm{b}} = \sqrt{\sum_{i=1}^{I} |(\mathbf{w}_{i,j-1}^{\mathrm{b}})^H \tilde{\mathbf{y}}_{i,j}^{\mathrm{b}}|^2}$
 $\varphi(\sigma_{\mathrm{s},j}^{\mathrm{b}}) = (\sigma_{\mathrm{s},j}^{\mathrm{b}})^{\beta-2}$
 $\bar{\mathbf{q}}_j^{\mathrm{b}} = \eta^{\mathrm{b}}\bar{\mathbf{q}}_{j-1}^{\mathrm{b}} + \frac{(1-\eta^{\mathrm{b}})\varphi(\sigma_{\mathrm{s},j}^{\mathrm{b}})}{I} \sum_{i=1}^{I} Y_{i,j}^* \mathbf{x}_{i,j}^{\mathrm{b}}$
 $\bar{\mathbf{R}}_j^{\mathrm{b}} = \eta^{\mathrm{b}}\bar{\mathbf{R}}_{j-1}^{\mathrm{b}} + \frac{(1-\eta^{\mathrm{b}})\varphi(\sigma_{\mathrm{s},j}^{\mathrm{b}})}{I} \sum_{i=1}^{I} \mathbf{x}_{i,j}^{\mathrm{b}} (\mathbf{x}_{i,j}^{\mathrm{b}})^H$
7: Update series expansion coefficients with (87)
 $\hat{\mathbf{b}}_j = [(\bar{\mathbf{R}}_j^{\mathrm{b}})^{-1}\bar{\mathbf{q}}_j^{\mathrm{b}}]^*$
8: Extract near-end signal
 $\hat{S}_{i,j} = (\mathbf{w}_j^{\mathrm{b}})^H \tilde{\mathbf{y}}_{i,j}^{\mathrm{b}}$
9: **end for**

After updating the CTF filter $\hat{\mathbf{a}}_{i,j}$, we use it to construct $\mathbf{x}_{i,j}^{\mathrm{b}}$. The solution of (73) is given by

$$\mathbf{w}_j^{\mathrm{b}} = \frac{\left(\bar{\mathbf{G}}_j^{\mathrm{b}}\right)^{-1} \mathbf{e}_{N+1}}{\mathbf{e}_{N+1}^T \left(\bar{\mathbf{G}}_j^{\mathrm{b}}\right)^{-1} \mathbf{e}_{N+1}}. \tag{81}$$

Following the previous analysis, one can express the inverse of $\bar{\mathbf{G}}_j^{\mathrm{b}}$ as

$$\left(\bar{\mathbf{G}}_j^{\mathrm{b}}\right)^{-1} = \begin{bmatrix} \left(\bar{S}_j^{\mathrm{b}}\right)^{-1} & -\left(\bar{S}_j^{\mathrm{b}}\right)^{-1} \left(\bar{\mathbf{c}}_j^{\mathrm{b}}\right)^H \\ -\left(\bar{S}_j^{\mathrm{b}}\right)^{-1} \bar{\mathbf{c}}_j^{\mathrm{b}} & \left(\bar{\mathbf{R}}_j^{\mathrm{b}}\right)^{-1} + \left(\bar{S}_j^{\mathrm{b}}\right)^{-1} \bar{\mathbf{D}}_j^{\mathrm{b}} \end{bmatrix}, \tag{82}$$

where

$$\bar{S}_j^{\mathrm{b}} = \left(\bar{\sigma}_{\mathrm{y},j}^{\mathrm{b}}\right)^2 - \left(\bar{\mathbf{q}}_j^{\mathrm{b}}\right)^H \left(\bar{\mathbf{R}}_j^{\mathrm{b}}\right)^{-1} \bar{\mathbf{q}}_j^{\mathrm{b}} \tag{83}$$

is the Schur complement of $\bar{\mathbf{R}}_j^{\mathrm{b}}$ in $\bar{\mathbf{G}}_j^{\mathrm{b}}$ and

$$\bar{\mathbf{c}}_j^{\mathrm{b}} = \left(\bar{\mathbf{R}}_j^{\mathrm{b}}\right)^{-1} \bar{\mathbf{q}}_j^{\mathrm{b}}, \tag{84}$$

$$\bar{\mathbf{D}}_j^{\mathrm{b}} = \bar{\mathbf{c}}_j^{\mathrm{b}} \left(\bar{\mathbf{c}}_j^{\mathrm{b}}\right)^H. \tag{85}$$

Algorithm 2: Alternating element-wise iterative source steering algorithm.

Setting forgetting factors η^{a} and η^{b}
  Initial parameters
  $\mathbf{X}_{i,0} = \mathbf{0}_{L\times N}$, $\hat{\mathbf{a}}_{i,0} = \mathbf{0}_{L\times 1}$, $\hat{\mathbf{b}}_0 = [1 \quad \mathbf{0}^T_{(N-1)\times 1}]^T$,
  $\mathbf{q}^{\mathrm{a}}_{i,0} = \mathbf{0}_{L\times 1}$, $\bar{\mathbf{q}}^{\mathrm{b}}_0 = \mathbf{0}_{N\times 1}$,
  $\mathbf{R}^{\mathrm{a}}_{i,0} = 10^{-4}\mathbf{I}_L$, $\bar{\mathbf{R}}^{\mathrm{b}}_0 = 10^{-4}\mathbf{I}_N$
for $j = 1; j < J; j = j + 1$ **do**
  Insert $X_{n,i,j}$ into $\mathbf{X}_{i,j-1}$ to get $\mathbf{X}_{i,j}$
  Update CTF filter associated statistics with (14), (40), (51), (52), (53), (54)
  for $l = 1; l < L; l = l + 1$ **do**
    Calculate the steering stepsize for the l-th element in the CTF filter with (92)
    $U^{\mathrm{a}}_{i,j,l} = \dfrac{(q^{\mathrm{a}}_{i,j,l})^* - \hat{\mathbf{a}}^T_{j-1}\mathbf{r}^{\mathrm{a}}_{i,j,l}}{R^{\mathrm{a}}_{i,j,l,l}}$
    Update the l-th element in the CTF filter with (88)
    $\hat{A}_{i,j,l} = \hat{A}_{i,j-1,l} + U^{\mathrm{a}}_{i,j,l}$
  end for
  Update series expansion coefficients associated statistics with (15), (58), (69), (70), (71), (72)
  for $n = 1; n < N; n = n + 1$ **do**
    Calculate the steering stepsize for the n-th series expansion coefficient with (97)
    $u^{\mathrm{b}}_{j,n} = \dfrac{(\bar{q}^{\mathrm{b}}_{j,n})^* - \hat{\mathbf{b}}^T_{j-1}\bar{\mathbf{r}}^{\mathrm{b}}_{j,n}}{\bar{r}^{\mathrm{b}}_{j,n,n}}$,
    Update the n-th series expansion coefficient with (93)
    $\hat{b}_{j,n} = \hat{b}_{j-1,n} + u^{\mathrm{b}}_{j,n}$
  end for
  Extract near-end signal with (63)
  $\hat{S}_{i,j} = (\mathbf{w}^{\mathrm{b}}_j)^H \tilde{\mathbf{y}}^{\mathrm{b}}_{i,j}$
end for

Substituting (82) into (81) gives

$$\left(\bar{S}^{\mathrm{b}}_j\right)^{-1}\begin{bmatrix} 1 \\ -\mathbf{b}^*_j \end{bmatrix} = \begin{bmatrix} \left(\bar{S}^{\mathrm{b}}_j\right)^{-1} \\ -\left(\bar{S}^{\mathrm{b}}_j\right)^{-1}\bar{\mathbf{c}}^{\mathrm{b}}_j \end{bmatrix}. \tag{86}$$

Therefore, the CTF filter can be updated as

$$\hat{\mathbf{b}}_j = \left[\left(\bar{\mathbf{R}}^{\mathrm{b}}_j\right)^{-1}\bar{\mathbf{q}}^{\mathrm{b}}_j\right]^*. \tag{87}$$

We summarize the AIP algorithm in Algorithm 1. Based on (80) and (87), RLS-like algorithms can be derived, which, however, will be left to the reader's investigation.

B. Alternating Element-Wise Iterative Source Steering

In the following, we extend a previously proposed EISS [57] algorithm for the use in our bilinear alternating optimization. We first construct $\mathbf{x}^{\mathrm{a}}_{i,j}$ with $\hat{\mathbf{b}}_{j-1}$ and update $\hat{\mathbf{a}}_{i,j}$. As we fix the first element in $\mathbf{w}^{\mathrm{a}}_{i,j}$ to 1, the original EISS can be further simplified by skipping calculating the first steering stepsize. Therefore, one can directly update every element in the CTF filter. The update rule is given by

$$\hat{A}_{i,j,l} = \hat{A}_{i,j-1,l} + U^{\mathrm{a}}_{i,j,l},\ l = 1, 2, \ldots, L, \tag{88}$$

where $\hat{A}_{i,j,l}$ is the estimate of $A_{i,j,l}$ and $U^{\mathrm{a}}_{i,j,l}$ is a steering stepsize yet to be determined. Substituting (88) into (48) gives

$$\mathcal{L}^{\mathrm{a},+} = \left[\mathbf{w}^{\mathrm{a}}_{i,j} - \tilde{\mathbf{u}}^{\mathrm{a}}_{i,j}\right]^H \mathbf{G}^{\mathrm{a}}_{i,j}\left[\mathbf{w}^{\mathrm{a}}_{i,j} - \tilde{\mathbf{u}}^{\mathrm{a}}_{i,j}\right], \tag{89}$$

where

$$\tilde{\mathbf{u}}^{\mathrm{a}}_{i,j} = \begin{bmatrix} 0 & \left(U^{\mathrm{a}}_{i,j,1}\right)^* & \ldots & \left(U^{\mathrm{a}}_{i,j,L}\right)^* \end{bmatrix}^T. \tag{90}$$

Forcing the derivative of (89) with respect to $(U^{\mathrm{a}}_{i,j,l})^*$ to be 0, one can determine the steering stepsize as

$$U^{\mathrm{a}}_{i,j,l} = \frac{\left(\mathbf{w}^{\mathrm{a}}_{i,j-1}\right)^H \mathbf{g}^{\mathrm{a}}_{i,j,l+1}}{G^{\mathrm{a}}_{i,j,l+1,l+1}}, \tag{91}$$

where $\mathbf{g}^{\mathrm{a}}_{i,j,l+1}$ denotes the $(l+1)$th column of $\mathbf{G}^{\mathrm{a}}_{i,j}$ and $g^{\mathrm{a}}_{i,j,l+1,l+1}$ is the $(l+1)$th diagonal element of $\mathbf{G}^{\mathrm{a}}_{i,j}$. Following the structure of (44) and (49), we have

$$U^{\mathrm{a}}_{i,j,l} = \frac{\left(q^{\mathrm{a}}_{i,j,l}\right)^* - \hat{\mathbf{a}}^T_{j-1}\mathbf{r}^{\mathrm{a}}_{i,j,l}}{R^{\mathrm{a}}_{i,j,l,l}}, \tag{92}$$

where $\mathbf{r}^{\mathrm{a}}_{i,j,l}$ denotes the lth column of $\mathbf{R}^{\mathrm{a}}_{i,j}$, $q^{\mathrm{a}}_{i,j,l}$ is the lth element of $\mathbf{q}^{\mathrm{a}}_{i,j}$ and $R^{\mathrm{a}}_{i,j,l,l}$ is the l-th diagonal element of $\mathbf{R}^{\mathrm{a}}_{i,j}$. In comparison with (92), (91) needs to update $\mathbf{R}^{\mathrm{a}}_{i,j}$ rather than $\mathbf{G}^{\mathrm{a}}_{i,j}$, which helps reduce the computation cost. Then the CTF filter $\hat{\mathbf{a}}_{i,j}$ is updated with (88).

Note that the coefficients b_n are independent of frequency. Therefore, we use lower case to denote the associated scalars. Now, we construct $\mathbf{x}^{\mathrm{b}}_{i,j}$ with the estimated $\hat{\mathbf{a}}_{i,j}$ and update $\hat{\mathbf{b}}_j$. Similarly, the nonlinear coefficients are updated as

$$\hat{b}_{j,n} = \hat{b}_{j-1,n} + u^{\mathrm{b}}_{j,n},\ n = 1, 2, \ldots, N, \tag{93}$$

where $\hat{b}_{j,n}$ is the estimate of b_n in time frame j and $u^{\mathrm{b}}_{j,n}$ is a steering stepsize to be determined. Substituting (93) into (66) gives the following expression:

$$\mathcal{L}^{\mathrm{b},+} = \left(\mathbf{w}^{\mathrm{b}}_j - \tilde{\mathbf{u}}^{\mathrm{b}}_j\right)^H \bar{\mathbf{G}}^{\mathrm{b}}_j\left(\mathbf{w}^{\mathrm{b}}_j - \tilde{\mathbf{u}}^{\mathrm{b}}_j\right), \tag{94}$$

where

$$\tilde{\mathbf{u}}^{\mathrm{b}}_j = \begin{bmatrix} 0 & \left(u^{\mathrm{b}}_{j,1}\right)^* & \ldots & \left(u^{\mathrm{b}}_{j,N}\right)^* \end{bmatrix}^T. \tag{95}$$

By optimizing (94) with respect to $(u^{\mathrm{b}}_{j,n})^*$, the steering stepsize is determined as

$$u^{\mathrm{b}}_{j,n} = \frac{\left(\mathbf{w}^{\mathrm{b}}_{j-1}\right)^H \bar{\mathbf{g}}^{\mathrm{b}}_{j,n+1}}{\bar{g}^{\mathrm{b}}_{j,n+1,n+1}}, \tag{96}$$

where $\bar{\mathbf{g}}^{\mathrm{b}}_{j,n+1}$ is the $(n+1)$th column of $\bar{\mathbf{G}}^{\mathrm{b}}_j$ and $\bar{g}^{\mathrm{b}}_{j,n+1,n+1}$ is the $(n+1)$th diagonal element of $\bar{\mathbf{G}}^{\mathrm{b}}_j$. By considering the structure of (44) and (49), we have

$$u^{\mathrm{b}}_{j,n} = \frac{\left(\bar{q}^{\mathrm{b}}_{j,n}\right)^* - \hat{\mathbf{b}}^T_{j-1}\bar{\mathbf{r}}^{\mathrm{b}}_{j,n}}{\bar{r}^{\mathrm{b}}_{j,n,n}}, \tag{97}$$

where $\bar{\mathbf{r}}^{\mathrm{b}}_{j,n}$, $\bar{q}^{\mathrm{b}}_{j,n}$ and $\bar{r}^{\mathrm{b}}_{j,n,n}$ are the nth column of $\bar{\mathbf{R}}^{\mathrm{b}}_j$, nth element of $\bar{\mathbf{q}}^{\mathrm{b}}_j$ and n-th diagonal element of $\bar{\mathbf{R}}^{\mathrm{b}}_j$, respectively. Then, the nonlinear coefficients $\hat{\mathbf{b}}_j$ are updated using (93). The AEISS algorithm is summarized in Algorithm 2.

TABLE II
MAIN EQUATIONS AND DOMINANT COMPUTATIONAL COMPLEXITY FOR EACH STEP OF ALL SBSS-NAEC ALGORITHMS

	IP	EISS	AIP	AEISS
Step 1	Eq. (27)-(29) $\mathcal{O}\left[(NL+1)^2\right]$	Eq. (27)-(29) $\mathcal{O}\left[(NL+1)^2\right]$	Eq. (51)-(54), (69)-(72) $\max\left[\mathcal{O}\left(N^2\right),\mathcal{O}\left(L^2\right)\right]$	Eq. (51)-(54), (69)-(72) $\max\left[\mathcal{O}\left(N^2\right),\mathcal{O}\left(L^2\right)\right]$
Step 2	Eq. (33), (34) $\mathcal{O}\left[(NL+1)^3\right]$	Eq. (34), (35), (38) $\mathcal{O}\left[(NL+1)^2\right]$	Eq. (80), (87) $\max\left[\mathcal{O}\left(N^3\right),\mathcal{O}\left(L^3\right)\right]$	Eq. (88), (92), (93), (97) $\max\left[\mathcal{O}\left(N^2\right),\mathcal{O}\left(L^2\right)\right]$
Step 3	Eq. (24) $\mathcal{O}(NL)$	Eq. (24) $\mathcal{O}(NL)$	Eq. (63) $\max\left[\mathcal{O}(N),\mathcal{O}(L)\right]$	Eq. (63) $\max\left[\mathcal{O}(N),\mathcal{O}(L)\right]$
Overall	$\mathcal{O}\left[(NL+1)^3\right]$	$\mathcal{O}\left[(NL+1)^2\right]$	$\max\left[\mathcal{O}\left(N^3\right),\mathcal{O}\left(L^3\right)\right]$	$\max\left[\mathcal{O}\left(N^2\right),\mathcal{O}\left(L^2\right)\right]$

V. COMPLEXITY ANALYSIS

We now analyze the computational complexity of the proposed AIP and AEISS algorithms and compare them with the conventional IP and EISS algorithms. All algorithms consist of the following three fundamental steps: 1) updating the associated statistics, 2) updating the corresponding filters, and 3) extracting the near-end signal. The computational complexity depends on the nonlinear echo model, the filter length, and the chosen optimization method. The dominant complexity and the associated equation for each step to process a single time-frequency component is presented in Table II. For IP-based methods, the dominant computational cost comes from the inversion of the auxiliary covariance matrix. Therefore, the complexity of the original IP algorithm is $\mathcal{O}[(NL+1)^3]$. With the bilinear nonlinear echo model and one-step strategy, AIP successfully reduces it to $\max[\mathcal{O}(N^3),\mathcal{O}(L^3)]$. As for EISS-based methods, the computational cost is dominated by updating the associated statistics and by the calculation of the steering stepsizes. The complexity of the original EISS algorithm is $\mathcal{O}[(NL+1)^2]$ while AEISS reduces it to $\max[\mathcal{O}(N^2),\mathcal{O}(L^2)]$. It is clear, that the computational complexity of EISS-based methods is one order lower than the IP-based counterpart. Obviously, for large P and L, the AIP and AEISS algorithms are much more computationally efficient than the conventional counterparts, indicating that they are much more suitable for real-time communications (RTC) in low-resource hardware.

VI. SIMULATIONS AND RESULTS

For consistency, we use the same speech signals as in the paper [64]. The sampling rate of all signals is 16 kHz. We randomly picked two AIRs from RWCP_E2A [81], which correspond to a reverberation time of approximately 300 ms. To model loudspeaker distortions, we consider the hard-clipping [82] function, which is of the following form:

$$f\left[x(t)\right]=\begin{cases}-\rho, & x(t)<-\rho\\ x(t), & |x(t)|\le\rho\\ \rho, & x(t)>\rho\end{cases}, \tag{98}$$

where $\rho>0$ is the maximum output amplitude of the loudspeaker. In our simulation, we set the value of ρ consistent to that in the paper [64], i.e., $\rho=0.2x_{\max}$, where $x_{\max}=\max(|x(t)|)$ is the maximum amplitude of the original signal. To model the nonlinearity, we set the order of odd power series expansion to $N=5$. A von Hann window is used and the window length is 1024 samples. The overlap between consecutive frames is 75%. Since the STFT window is much shorter than the AIR, we set the CTF filter length to $L=5$. All experiments are conducted on a laptop with i7-12700H CPU. For the IP and EISS methods, we set the forgetting factors to $\eta^{\mathrm{m}}=0.992$. To make the algorithms robust, the auxiliary matrix $\tilde{\mathbf{G}}^{\mathrm{m}}_{i,0}$ is set to $10^{-3}\times\mathbf{I}_{NL+1}$. For the AIP and AEISS methods, we set $\eta^{\mathrm{a}}=\eta^{\mathrm{b}}=0.98$. The auxiliary matrices $\mathbf{G}^{\mathrm{a}}_{i,0}$ and $\bar{\mathbf{G}}^{\mathrm{b}}_{0}$ are set to be $10^{-4}\times\mathbf{I}_L$ and $10^{-4}\times\mathbf{I}_N$, respectively, and the two auxiliary vectors $\mathbf{q}^{\mathrm{a}}_{i,0}$ and $\mathbf{q}^{\mathrm{b}}_{0}$ are set to be $\mathbf{0}_{L\times1}$ and $\mathbf{0}_{N\times1}$, respectively. The CTF filter and nonlinear coefficients are initialized as $\hat{\mathbf{a}}_{i,0}=\mathbf{0}_{L\times1}$ and $\hat{\mathbf{b}}_0=[1\quad\mathbf{0}^T_{(N-1)\times1}]^T$. The shape parameter β is set to $\beta=0.4$ to be consistent with [64]. Note that with this configuration, AIP and AEISS algorithms have similar steady-state performance as their conventional counterparts as illustrated in Section VI-A. To evaluate their performance, we use the echo return loss enhancement (ERLE) [30], [83] for the single-talk case and the true echo return loss enhancement (tERLE) [30], [38] for the double-talk case. The implementation of these two measures involves smoothing, utilizing samples within 0.2 s vicinity. Besides, an experiment with real recordings is carried out where the Perceptual Evaluation of Speech Quality (PESQ) [84] and the Short-Time Objective Intelligibility (STOI) [85] metrics are used to measure the quality of the obtained near-end signals.

We compare the proposed algorithms with both the original SBSS-NAEC algorithms and a single-channel variant of [31], namely, a state-of-the-art state-space model-based NAEC (SSM-NAEC) algorithm.

A. Tracking Ability

A 20-second-long AR(1) signal with $x_{\max}$=1, generated by filtering a white Gaussian noise with the system $1/(1-0.8z^{-1})$,

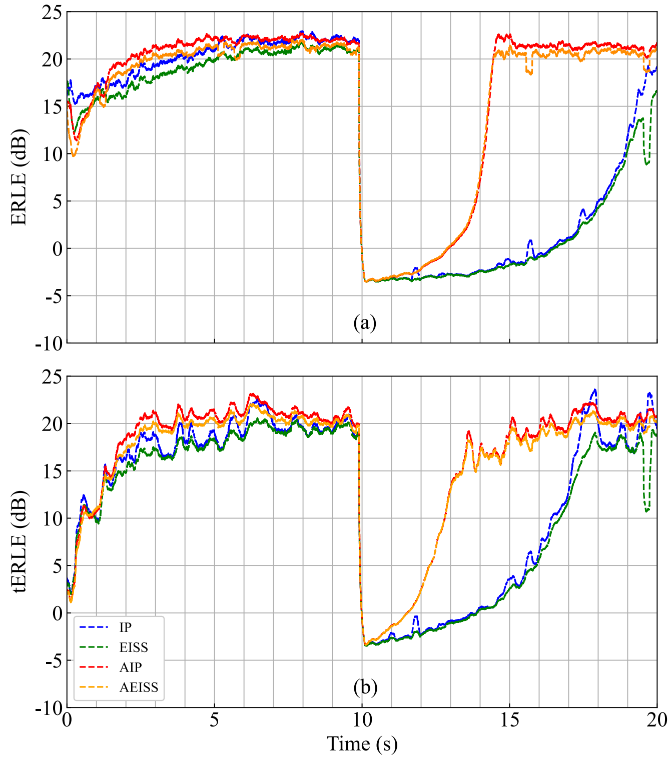


Fig. 3. Tracking ability of all the compared SBSS-NAEC algorithms. The far-end signal is an AR(1) signal. The AIR changes at the 10 s. (a) ERLE over time in the single-talk case and (b) tERLE over time in double-talk.

is utilized as the far-end signal. Subsequently, it undergoes the hard clipping function to produce the nonlinear loudspeaker signal. For the first 10 seconds, the nonlinearly distorted signal is convolved with the first AIR and it is then convolved with the second AIR. In other words, the AIR changes at $t=10$ s. We consider both single-talk and double-talk situations. For the double-talk case, the female speech signal from [64] is used as near-end signal and the corresponding signal-to-echo ratio (SER) is set to 0 dB. A white Gaussian noise is added as background noise corresponding with a signal-to-noise ratio (SNR) of 60 dB. Please note that we refrain from comparing the tracking capabilities with SSM-NAEC, as it cannot attain AEC performance comparable to SBSS-NAEC algorithms during double-talk scenarios. This will be substantiated in Section (VI-C) and Section (VI-D). As seen in Fig. 3, AIP and AEISS algorithms achieved similar steady-state performance with their conventional counterparts in the first 10 seconds. For the second 10 seconds, the proposed AIP and AEISS have significantly better tracking ability since they have less number of free parameters to estimate.

B. Performance in Single-Talk Case

Now, we compare the performance of all SBSS-NAEC algorithms during single talk using the data from [64]. A 10-second long male speech signal is used as the far-end signal. This signal is passed through the hard clipping function and is then convolved with the first AIR to generate the nonlinear echo. White Gaussian noise is added as background noise with an SNR of 60 dB. The ERLE performance achieved by all compared algorithms are plotted in Fig. 4. As observed, SSM-NAEC demonstrates performance akin to that of IP and EISS. Furthermore, in the single-talk scenario, the AIP and AEISS algorithms exhibit markedly superior performance.

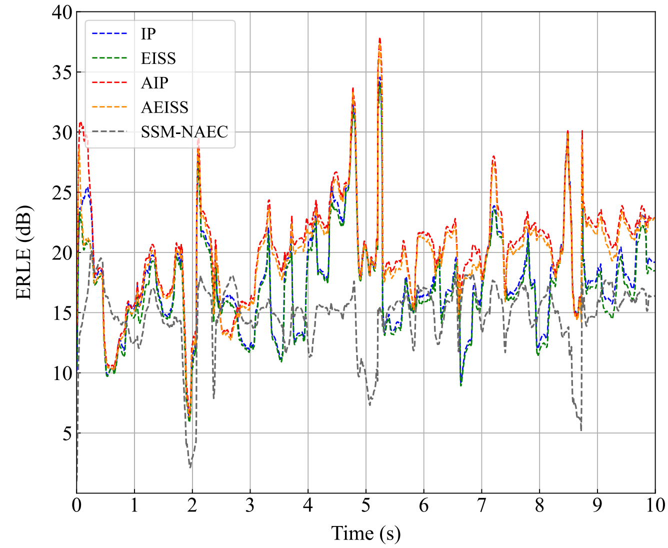


Fig. 4. ERLE performance achieved by all compared algorithms in a single-talk situation.

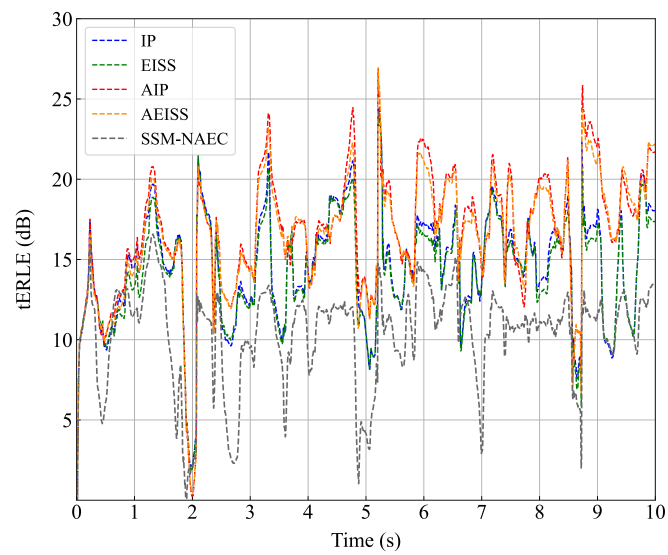


Fig. 5. tERLE performance achieved by all compared algorithms in a double-talk situation.

C. Performance in Double-Talk Case

We now compare the performance of all compared algorithms in the double-talk case, again, using the data from [64]. The nonlinear echo is generated following the similar process as in Section VI-B. A female speech signal of 10-second long is used as the near-end signal with an SER of 0 dB. White Gaussian noise is added as background noise with an SNR of 60 dB. In Fig. 5, we plot the tERLE performance of all compared algorithms. Clearly, the SSM-NAEC algorithm falls short of achieving comparable AEC performance when compared to SBSS-NAEC algorithms

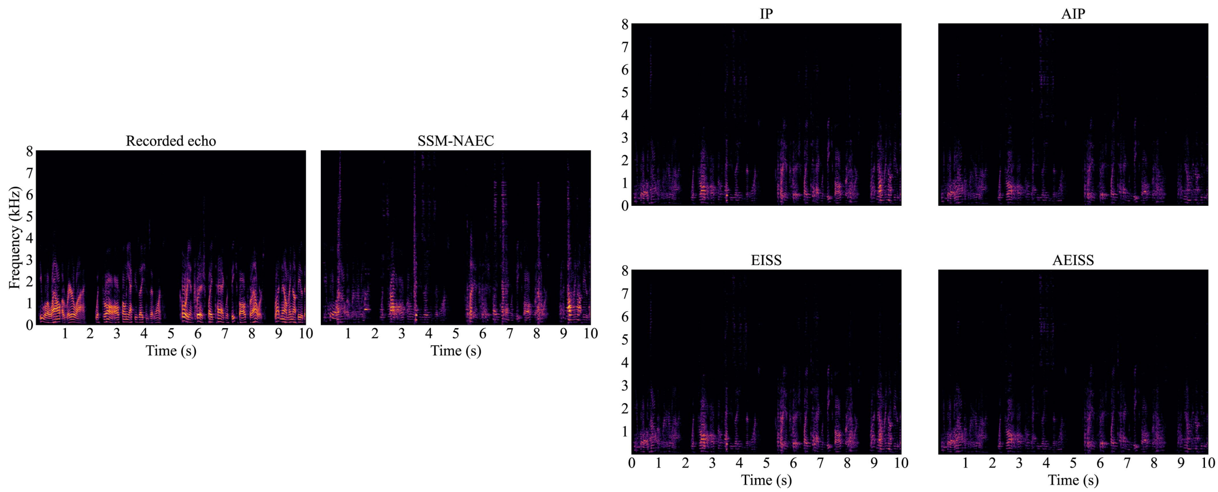


Fig. 6. Spectrograms of the recorded echo and error signals generated by all compared algorithms.

in double-talk scenarios. Additionally, the AIP and AEISS algorithms outperform IP and EISS, underscoring their superiority in addressing double-talk situations in AEC applications.

D. Experiment With Real Recordings

Now, we compare the AEC performance with real recorded echoes. The aforementioned 10-second speech signal from a male speaker is played by a low-cost loudspeaker and picked up by a mobile phone in an office environment, sampled at 16 kHz. Subsequently, a 10-second speech signal from a female speaker is introduced as the near-end signal at an SER of 0 dB. It is important to note that we refrain from adding additional background noise, as the office noise is already captured in the recordings. We define the error signal as

$$e(t) = s(t) - \hat{s}(t). \tag{99}$$

The spectrograms of the error signals generated by all compared algorithms are shown in Fig. 6. To enhance clarity, the lower bound in the figure is set to 55 dB below the highest power in the data. As seen, the error signal generated by SSM-NAEC has much higher power than those generated by the SBSS-NAEC algorithms. Besides, SSM-NAEC generates some new components in high frequency bins, indicating that it caused distortion. Moreover, it is also observed that the power of the error signals generated by AIP and AEISS is lower than those generated by their conventional counterparts. The obtained near-end signals are also evaluated with PESQ and STOI, which are shown in Table III. Remarkably, the proposed AIP and AEISS algorithms achieve notably higher PESQ and STOI values compared to other methods, underscoring their ability to obtain high-quality near-end signals.

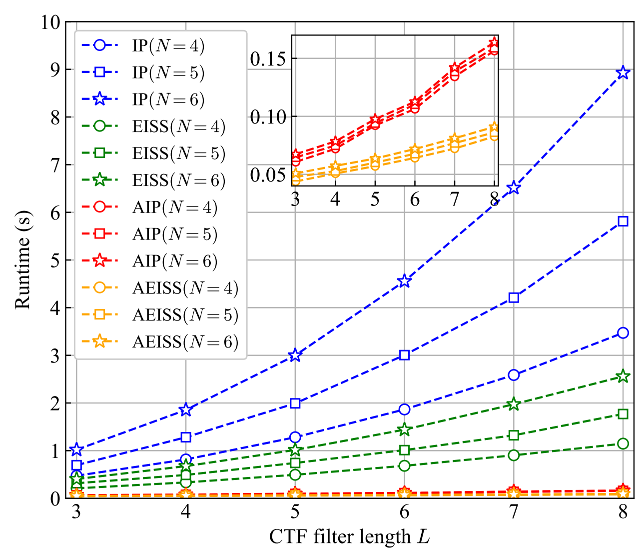


Fig. 7. The average runtime to process 1-second long signal with a 16 kHz sampling rate.

TABLE III
PESQ AND STOI OF THE OBTAINED NEAR-END SIGNALS WITH REAL RECORDINGS

Algorithms	PESQ	STOI
Unprocessed	1.22	0.73
SSM-NAEC	1.51	0.84
IP	1.81	0.92
EISS	1.77	0.91
AIP	2.15	0.95
AEISS	2.09	0.95

E. Runtime Comparison

Finally, we compare the runtime of SBSS-NAEC algorithms. Two 20-minute long white Gaussian noise signals are used as the far-end and near-end signals. We compare the computation time of the studied algorithms, which covers all the aforementioned three steps. The odd power series expansion order N is set, respectively, to 4, 5, and 6 while the CTF filter length L varies from 3 to 8. The average runtime for processing a 1-second long signal with all studied algorithms is plotted in Fig. 7. As seen, the runtime of all SBSS-NAEC algorithms increases with the the value of of N and L. Under the same configuration, EISS-based methods have less runtime than the IP-based methods. The proposed AIP and AEISS algorithms are much more efficient than the conventional IP and EISS algorithms, and the difference is more significant as the value of L and N increases. This validates another property of the proposed AIP and AEISS algorithms, i.e., besides being able to achieve better performance, it also has lower computational complexity as compared to their conventional counterparts.

VII. Conclusion

In this paper, we adopted a bilinear model to represent nonlinear echoes. To estimate the series expansion coefficients and the CTF filter in this model, we presented a bilinear alternating optimization framework. Under this framework, two algorithms, i.e., the AIP and AEISS, were derived, both exploit the independence criteria to estimate the model parameters. We showed that the proposed AIP and AEISS algorithms are capable to achieve nonlinear AEC in both the single-talk and double-talk scenarios. Since the bilinear representation consists of less parameters compared to a conventional CTF model, the developed AIP and AEISS algorithms have demonstrated interesting properties as compared to the conventional IP and EISS algorithms including improved NAEC performance, better tracking ability, and reduced complexity.

Appendix A
Derivation of the Auxiliary Function

With the definition of

$$\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}} = \sqrt{\sum_{i=1}^{I} \left| \left(\mathbf{w}_{i,j}^{\mathrm{m}}\right)^H \tilde{\mathbf{y}}_{i,j}^{\mathrm{m}} \right|^2}, \tag{100}$$

the log likelihood function can be expressed as

$$\mathcal{F}\left(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}}\right) = -\log p(\mathbf{s}_j). \tag{101}$$

As $\mathbf{s}_j$ follows a super Gaussian distribution, the following auxiliary function can be used [48]

$$\mathcal{F}^{+}\left(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}}\right) = \frac{\mathcal{F}'\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right)}{2\sigma_{\mathrm{s},j}^{\mathrm{m}}} \left(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}}\right)^2 + \mathcal{F}\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right) - \frac{\sigma_{\mathrm{s},j}^{\mathrm{m}} \mathcal{F}'\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right)}{2}, \tag{102}$$

where $(\cdot)'$ represents the derivative and $\sigma_{i,j}^{\mathrm{m}}$ is defined in (29). The above auxiliary function satisfies

$$\mathcal{F}^{+}\left(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}}\right) \geq \mathcal{F}\left(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}}\right). \tag{103}$$

with equality if and only if $\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}} = \sigma_{\mathrm{s},j}^{\mathrm{m}}$, i.e., $\mathbf{w}_{i,j}^{\mathrm{m}} = \mathbf{w}_{i,j-1}^{\mathrm{m}}$. Therefore, instead of minimizing $\mathcal{F}(\tilde{\sigma}_{i,j}^{\mathrm{m}})$ at each time frame, one can minimize $\mathcal{F}^{+}(\tilde{\sigma}_{\mathrm{s},j}^{\mathrm{m}})$. Note that

$$\frac{\mathcal{F}'\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right)}{2\sigma_{\mathrm{s},j}^{\mathrm{m}}} = \phi\left(\sigma_{\mathrm{s},j}^{\mathrm{m}}\right), \tag{104}$$

which is defined in (28). Now substituting (100) and (103) into (25) and considering that $\det \mathbf{W}_{i,j}^{\mathrm{m}} = 1$, we obtain (26), where we neglected irrelevant constant terms.

References

[1] A. Gilloire, E. Moulines, D. Slock, and P. Duhamel, "State of the art in acoustic echo cancellation," in *Digital Signal Processing in Telecommunications*, A. R. Figueiras-Vidal, Eds., Berlin, Germany: Springer, 1996, pp. 45–91.

[2] S. L. Gay and J. Benesty Eds., *Acoustic Signal Processing for Telecommunication*. Boston, MA, USA: Kluwer, 2000.

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

[4] S. L. Gay, "Acoustic echo cancellation for duplex audio communication," *J. Acoust. Soc. Amer.*, vol. 99, no. 4, pp. 2504–2529, 1996.

[5] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*. New York, NY, USA: Springer, 2007.

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

[7] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, Oct. 2018.

[8] X. Wang, G. Huang, J. Benesty, J. Chen, and I. Cohen, "Time difference of arrival estimation based on a Kronecker product decomposition," *IEEE Signal Process. Lett.*, vol. 28, pp. 51–55, 2021.

[9] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[10] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ, USA: Wiley, 2003.

[11] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ, USA: Wiley, 2008.

[12] C.-L. Stanciu, J. Benesty, C. Paleologu, R.-L. Costea, L.-M. Dogariu, and S. Ciochină, "Decomposition-based wiener filter using the Kronecker product and conjugate gradient method," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 124–138, 2024.

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

[14] 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, Sep. 1989.

[15] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3125–3128.

[16] G. Su, J. Jin, Y. Gu, and J. Wang, "Performance analysis of ℓ_0-norm constraint least mean square algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2223–2235, May 2012.

[17] A. E. Albert and L. S. Gnrdner Jr., *Stochatic Approxinmtion and Nodinem Regreaion*. Cambridge, MA, USA: MIT Press, 1967.

[18] 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.

[19] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[20] C. Paleologu, J. Benesty, and S. Ciochină, "An improved proportionate NLMS algorithm based on the ℓ_0 norm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 309–312.

[21] V. Panuska, "An adaptive recursive-least-squares identification algorithm," in *Proc. IEEE Symp. Adaptive Process. Decis. Control*, 1969, pp. 65–65.

[22] 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, 2008.

[23] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 5, pp. 903–918, May 2019.

[24] L.-M. Dogariu, J. Benesty, C. Paleologu, and S. Ciochină, "Identification of room acoustic impulse responses via Kronecker product decompositions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2828–2841, 2022.

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

[26] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 8, pp. 1539–1549, Aug. 2013.

[27] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the optimal and simplified Kalman filters for echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 580–584.

[28] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech, Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

[29] H. Buchner, J. Benesty, T. Gansler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1633–1644, Sep. 2006.

[30] 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, Sep. 2012.

[31] J. Park and J.-H. Chang, "State-space microphone array nonlinear acoustic echo cancellation using multi-microphone nearend speech covariance," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 10, pp. 1520–1534, Oct. 2019.

[32] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[33] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2001.

[34] S. Makino, *Audio Source Separation*. Berlin, Germany: Springer, 2018.

[35] D. W. E. L. Schobben and C. W. Sommen, "A frequency domain blind signal separation method based on decorrelation," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1855–1865, Aug. 2002.

[36] H. Buchner and W. Kellermann, "A fundamental relation between blind and supervised adaptive filtering illustrated for blind source separation and acoustic echo cancellation," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, 2008, pp. 17–20.

[37] T. S. Wada and B.-H. Juang, "Acoustic echo cancellation based on independent component analysis and integrated residual echo enhancement," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2009, pp. 205–208.

[38] 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.

[39] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.

[40] Y. Avargel and I. Cohen, "System identification in the short time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[41] 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.

[42] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1420–1434, Sep. 2009.

[43] 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 Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[44] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1988, pp. 2570–2573.

[45] S. L. Gay and R. J. Mammone, "Fast converging subband acoustic echo cancellation using RAP on the WE DSP16A," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, pp. 1141–1144.

[46] A. Yeredor, "On hybrid exact-approximate joint diagonalization," in *Proc. Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2009, pp. 312–315.

[47] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2011, pp. 189–192.

[48] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2012, pp. 1–4.

[49] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 2016.

[50] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Blind speech extraction based on rank-constrained spatial covariance matrix estimation with multivariate generalized Gaussian distribution," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1948–1963, 2020.

[51] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 236–240.

[52] R. Scheibler, "Independent vector analysis via log-quadratically penalized quadratic minimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 2509–2524, 2021.

[53] T. Nakashima and N. Ono, "Inverse-free online independent vector analysis with flexible iterative source steering," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2022, pp. 749–753.

[54] T. Nakashima, R. Ikeshita, N. Ono, S. Araki, and T. Nakatani, "Fast online source steering algorithm for tracking single moving source using online independent vector analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[55] Z. 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 Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 141–145.

[56] 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, 2021.

[57] 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," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2024, pp. 756–760.

[58] T. Koh and E. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 6, pp. 1445–1455, Dec. 1985.

[59] L. Tan and J. Jiang, "Adaptive Volterra filters for active control of nonlinear noise processes," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1667–1676, Aug. 2001.

[60] A. Guerin, G. Faucon, and R. L. Bouquin-Jeannes, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 672–683, Nov. 2003.

[61] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, no. 9, pp. 1747–1760, Sep. 2000.

[62] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 105–108.

[63] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Process.*, vol. 86, no. 6, pp. 1168–1181, Jun. 2006.

[64] 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. Amer.*, vol. 153, no. 1, pp. 88–95, Jan. 2023.

[65] J. Benesty, C. Paleologu, and S. Ciochină, "On the identification of bilinear forms with the Wiener filter," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 653–657, May 2017.

[66] L. Dogariu, C. Paleologu, S. Ciochină, J. Benesty, and P. Piantanida, "Identification of bilinear forms with the Kalman filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 4134–4138.

[67] M. Z. Ikram, "Non-linear acoustic echo cancellation using cascaded Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 1320–1324.

[68] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statistician.*, vol. 58, no. 1, pp. 30–37, 2004.

[69] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory*. New York, NY, USA: Wiley, 2002.

[70] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[71] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge. U.K.: Cambridge Univ. Press, 2010.

[72] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[73] X. Wang, J. Benesty, G. Huang, and J. Chen, "A minimum variance distortionless response spectral estimator with Kronecker product filters," in *Proc. 30th Eur. Signal Process Conf.*, 2022, pp. 2261–2265.

[74] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[75] J. Dmochowski, J. Benesty, and S. Affes, "Linearly constrained minimum variance source localization and spectral estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1490–1502, Nov. 2008.

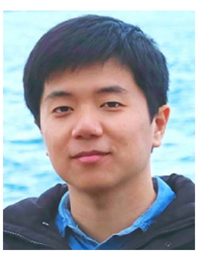
[76] K. M. Buckley and L. J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 311–319, Mar. 1986.

[77] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Rev.*, vol. 23, no. 1, pp. 53–60, Jan. 1981.

[78] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[79] F. Zhang, *The Schur Complement and Its Applications*. New York, NY, USA: Springer, 2005.
[80] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: An ASR corpus based on public domain audio books,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 5206–5210.
[81] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical sound database in real environments for sound sceneunderstanding and hands-free speech recognition,” in *Proc. Lang. Resour. Eval.*, 2000, pp. 965–0968.
[82] R. Cutler et al., “ICASSP 2022 acoustic echo cancellation challenge,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 9107–9111.
[83] M. Zeller and W. Kellermann, “Fast and robust adaptation of DFT-domain Volterra filters in diagonal coordinates using iterated coefficient updates,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1589–1604, Mar. 2010.
[84] “Mapping Function for transforming raw results scores to MOS-LQO,” ITU-T Rec. P. 862.1, 2003.
[85] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, pp. 2125–2136, Sep. 2011.

Xianrui Wang (Graduate Student Member, IEEE) received the bachelor’s degree in control science and technology in 2019 from Northwestern Polytechnical University, Xi’an, China, where he is currently working toward the Ph.D. degree in information and communication engineering with the Center of Intelligent Acoustics and Immersive Communications. He is currently a Visiting Research Fellow with Waseda University, Kitakyushu, Japan. His research interests include blind and semi-blind source separation and extraction, acoustic echo cancellation, speech dereverberation, and source localization. He is an active Reviewer of IEEE SIGNAL PROCESSING LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, *Sensors and Actuators A: Physical*, and IEEE ICASSP.



Yichen Yang (Graduate Student Member, IEEE) received the bachelor’s and master’s degrees from Northwestern Polytechnical University, Xi’an, China, in 2018 and 2021, respectively. He is currently working toward the Ph.D. degree in information and communication engineering with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi’an, China. He is also a Visiting Ph.D. student with Waseda University, Kitakyushu, Japan. His research interests include array signal processing, speech enhancement, blind source separation, and machine learning.

Andreas Brendel (Member, IEEE) received the B.Sc. and M.Sc. (with distinction) degrees in electrical engineering from the Friedrich-Alexander Universität Erlangen-Nürnberg, Germany, in 2014 and 2016, respectively, and the Dr.-Ing. (with distinction) degree from Friedrich-Alexander Universität Erlangen-Nürnberg, in 2022. From 2016 to 2022, he was with the Chair of multimedia communications and signal processing with Friedrich-Alexander Universität Erlangen-Nürnberg. He was a Visiting Scientist with Tsukuba University, Tsukuba, Japan from September to November 2016, and with Bar Ilan University, Ramat Gan, Israel, from Sepember to October 2018. In 2022, he joined the Fraunhofer Institute for Integrated Circuits IIS in Erlangen as an Associated Researcher. He authored or coauthored more than 50 peer-reviewed papers in journals and conference proceedings as well as several patents. His research interests include acoustic source separation and deep generative modeling for speech synthesis and coding.

Tetsuya Ueda (Student Member, IEEE) received the B.Sc. and M.E. degrees in information engineering and engineering from the University of Tsukuba, Tsukuba, Japan, in 2020 and 2022, respectively. He is currently working toward the Ph.D. degree with Waseda University, Kitakyushu, Japan. His research interests include acoustic signal processing, speech enhancement, and dereverberation.

Shoji Makino (Life Fellow, IEEE) received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively. In 1981, he joined NTT and the University of Tsukuba, Ibaraki, Japan, in 2009. He is currently a Professor with Waseda University, Kitakyushu, Japan. He has authored or coauthored of more than 400 articles in journals and conference proceedings and is responsible for more than 200 patents. His research interests include adaptive filtering technologies, blind source separation of convolutive mixtures of speech, the realization of acoustic echo cancellation, and acoustic signal processing for speech and audio applications. He was the recipient of the IEEE Signal Processing Society Leo L. Beranek Meritorious Service Award in 2022, ICA Unsupervised Learning Pioneer Award in 2006, IEEE MLSP Competition Award in 2007, IEEE SPS Best Paper Award in 2014, Achievement Award for Science and Technology from the Japanese Government in 2015, Hoko Award of the Hattori Hokokai Foundation in 2018, Honorary Member Award of the IEICE in 2022, Outstanding Contribution Award of the IEICE in 2018, Technical Achievement Award of the IEICE in 2017 and 1997, Outstanding Technological Development Award of the ASJ in 1995, and eight best paper awards. He was a member of the IEEE Jack S. Kilby Signal Processing Medal Committee during 2015–2018, and the James L. Flanagan Speech & Audio Processing Award Committee during 2008–2011. He was on IEEE SPS Board of Governors during 2018–2020, Technical Directions Board during 2013–2014, Awards Board during 2006–2008, Conference Board during 2002–2004, and a Fellow Evaluation Committee during 2018–2020. He was a Keynote Speaker at ICA 2007, Tutorial Speaker at ICASSP 2007, Interspeech 2011, and EMBC 2013. He was an Associate Editor for IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING during 2002–2005 and an Associate Editor for the *EURASIP Journal on Advances in Signal Processing* during 2005–2012. He was the Guest Editor of the Special Issue of the *IEEE Signal Processing Magazine* during 2013–2014. He was the Chair of SPS Audio and Acoustic Signal Processing Technical Committee during 2013–2014, and the Chair of the Blind Signal Processing Technical Committee of the IEEE Circuits and Systems Society during 2009–2010. He was the General Chair of IWAENC 2018, WASPAA 2007, IWAENC 2003, Organizing Chair of ICA 2003, and is the designated Plenary Chair of ICASSP 2012. Dr. Makino is an IEEE SPS Distinguished Lecturer during 2009–2010, an IEICE Fellow, a Board member of the ASJ, and a member of EURASIP.

Jacob Benesty received the master’s degree in microwaves from Pierre & Marie Curie University, France, in 1987, and the Ph.D. degree in control and signal processing from Paris-Saclay University, Paris, France, in April 1991, and the Honorary Doctorate (Doctor Technices Honoris Causa) degree from Aalborg University, Aalborg, in 2023, for his distinguished efforts in audio and acoustic signal processing. During his Ph.D. degree from 1989 to 1991, he worked on adaptive filters and fast algorithms with the Centre National d’Etudes des Telecommunications, Paris, France. From 1994 to 1995, he was with Telecom Paris University, France, on multichannel adaptive filters and acoustic echo cancellation. From 1995 to 2003, he was a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined the University of Quebec, INRS-EMT, Montreal, QC, Canada, as a Professor. He is also an Adjunct Professor with Aalborg University, Aalborg, Denmark, and a Guest Professor with Northwestern Polytechnical University, Xi’an, China. He has coauthored and coedited/coauthored numerous books in the area of acoustic signal processing. His research interests include signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead Researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. He is also conceived and designed the world- first PC-based multi-party hands-free full-duplex stereo conferencing system more than IP networks. He is the Editor of the book series Springer Topics in Signal Processing. He was the General Chair and Technical Chair of many international conferences and a member of several IEEE technical committees. Four of his journal papers were awarded by the IEEE Signal processing Society, in 2010. He was the recipient of the Gheorghe Cartianu Award from the Romanian Academy.

Walter Kellermann received the Dipl.-Ing. (Univ.) degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1983, and the Dr.-Ing. degree from the Technical University Darmstadt, Germany, in 1988. Since 1999, he has been a Professor of communications with the University of Erlangen-Nuremberg. From 1989 to 1990, he was a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ, USA. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany, to work on hands-free communication in cars. From 1993 to 1999, he was a Professor with the Fachhochschule Regensburg, where he also became the Director of the Institute of Applied Research in 1997. In 1999, he Co-founded DSP Solutions, a consulting firm in digital signal processing, and he joined the University Erlangen-Nuremberg, as a Professor and the Head of the Audio Research Laboratory. He was also a Visiting Fellow with Australian National University, Canberra, ACT, Australia, in 2016. He authored or coauthored more than 20 book chapters, more than 400 refereed papers in journals and conference proceedings, and more than 80 patents. His research interests include speech signal processing, array signal processing and machine learning, especially for acoustic signal processing. He was an Associate Editor and Guest Editor to various journals, including the IEEE Transactions on Speech and Audio Processing from 2000 to 2004, *IEEE Signal Processing Magazine* in 2015, IEEE Journal on Special Topics in Signal Processing in 2019. He is an Associate Editor for *EURASIP Journal on Applied Signal Processing*. He was the General Chair of eight mostly IEEE-sponsored workshops and conferences. He was a Distinguished Lecturer of the IEEE Signal Processing Society (SPS) from 2007 to 2008. He was the Chair of the IEEE SPS Technical Committee for Audio and Acoustic Signal Processing from 2008 to 2010, a Member of the IEEE James L. Flanagan Award Committee from 2011 to 2014, a Member of the SPS Board of Governors from 2013 to 2015, Vice President Technical Directions of the IEEE Signal Processing Society from 2016 to 2018, Member of the SPS Nominations Appointments Committee from 2019 to 2022. He is currently a Member of the SPS Fellow Evaluation Committee and as the Chair of the EURASIP Fellow Evaluation Committee. He was the co-recipient of ten best paper awards. He was also the recipient of Julius von Haast Fellowship by the Royal Society of New Zealand in 2012 and the Group Technical Achievement Award of the European Association for Signal Processing (EURASIP) in 2015. He was elevated to EURASIP Fellow in 2021.

Jingdong Chen (Fellow, IEEE) received the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences, Beijing, China, in 1998. He is currently a Professor with Northwestern Polytechnical University, Xi'an, China. Prior to this position, he was with Bell Laboratories, Murray Hill, NJ, USA, WeVoice Inc., NJ, Griffith University, Brisbane, QLD, Australia, and Advanced Telecommunication Research Institute International, Kyoto, Japan, for more than a decade. He is a co-author of a paper for which C. Pan was the recipient of the IEEE R10 (Asia-Pacific Region) Distinguished Student Paper Award (first prize) in 2016. His research interests include array signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, signal separation, speech communication, and artificial intelligence.

He was an Associate Editor for IEEE Transactions on Audio, Speech, and Language Processing from 2008 to 2014, as a Technical Committee (TC) Member of the IEEE Signal Processing Society (SPS) TC on Audio and Electroacoustics from 2007 to 2009 and a member of the IEEE SPS TC on Audio and Acoustic Signal Processing from 2018 to 2021. He is currently the Chair of IEEE Xi'an Section. He was the General Co-Chair of ACM WUWNET 2018 and IWAENC 2016, Technical Program Chair of IEEE TENCON 2013, Technical Program Co-Chair of IEEE WASPAA 2009, IEEE ChinaSIP 2014, IEEE ICSPCC 2014, and IEEE ICSPCC 2015, and helped organize many other conferences.

Dr. Chen was the recipient of the 2008 Best Paper Award from the IEEE Signal Processing Society (with Benesty, Huang, and Doclo), Best Paper Award from the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2011 (with Benesty), Bell Labs Role Model Teamwork Award twice, in 2009 and 2007, respectively, NASA Tech Brief Award twice, in 2010 and 2009, respectively, and the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998. He was also the recipient of the Japan Trust International Research Grant from the Japan Key Technology Center in 1998 and the Distinguished Young Scientists Fund from the National Natural Science Foundation of China in 2014.