Multiplicative Update Reformulation of Online Independent Vector Analysis for Source-Dependent Forgetting Factor Control

Kaito Masuko, Taishi Nakashima, and Nobutaka Ono
Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo, Japan
masuko-kaito@ed.tmu.ac.jp, taishi@tmu.ac.jp, onono@tmu.ac.jp

***Abstract*—In this paper, we reformulate online Auxiliary-function-based Independent Vector Analysis (AuxIVA) by applying a multiplicative update, which enables source-dependent control of the forgetting factor. In conventional online AuxIVA, a well-studied real-time blind source separation (BSS) method, the weighted covariance matrix, which represents key signal statistics, is updated using recursive estimation with a scalar forgetting factor that serves as a uniform weight to all sources. This uniform weighting may not be optimal, particularly when some sources move while others remain stationary, as it risks forgetting important statistical information that should be retained. In contrast, our proposed method expresses key statistics as a weighted covariance matrix of separated sources, thereby allowing the forgetting factor to be represented as a diagonal matrix. This enables more adaptive and precise statistical estimation by selectively retaining or forgetting signal statistics for each source. Simulation experiments demonstrate that the proposed method improves real-time source separation performance compared to conventional approaches, particularly in dynamic environments such as when sound sources move.**

***Index Terms*—multiplicative update, forgetting factor, online blind source separation, independent vector analysis**

I. Introduction

Blind source separation (BSS) is a signal processing technique for estimating source signals using only mixture signals observed at microphones, without requiring prior information such as the direction of signal arrival or the location of the microphone. Independent Vector Analysis (IVA) [1], [2], one of the widely studied BSS methods, has been investigated. Real-time extensions of BSS have also been extensively studied in response to the demand for real-time acoustic applications such as automatic speech recognition and hearing aid systems [3], [4]. In particular, online AuxIVA [5], an online variant of Auxiliary-function-based IVA (AuxIVA) [6], has gained attention due to its stability and computational efficiency. Online AuxIVA has seen many developments in recent years, including low latency processing [7]–[9], joint optimization with other processing such as dereverberation [10]–[12], fast update of the steering vectors [13], [14], and robustness to microphone rotation [15].

Online AuxIVA sequentially estimates the weighted covariance matrix by recursively averaging the past estimate and the instantaneous one in the current frame using a parameter called the forgetting factor. This parameter controls how much past information is retained at each update. In stationary environments, the covariance statistics remain time-invariant, making it important to retain them. In contrast, in dynamic environments, such as when sound sources move, past covariance statistics become outdated, and rapid forgetting is necessary to ensure accurate separation. Thus, dynamically controlling the forgetting factor is crucial for maintaining separation performance in real-world scenarios.

However, conventional online AuxIVA employs a scalar forgetting factor, which applies the same weight to all sources. As a result, important statistical information that should be retained may be lost, particularly when some sources move while others remain stationary. In such cases, the optimal approach would be to selectively forget only the moving source while preserving the statistics of stationary sources.

To address this limitation, we reformulate the demixing matrix update in online AuxIVA by applying a *multiplicative update*, which enables source-dependent control of the forgetting factor. Our proposed method expresses key statistics as a weighted covariance matrix of separated sources, thereby allowing the forgetting factor to be represented as a diagonal matrix. This enables more adaptive and precise statistical estimation by selectively retaining or forgetting signal statistics for each source. Simulation experiments confirm that the proposed method improves real-time source separation performance compared to conventional approaches, particularly in dynamic environments.

II. Problem Formulation

Let K be the number of microphones and sources, namely we consider the determined case. In the time-frequency domain, when the source signals are defined as $\boldsymbol{s}_{f,t} = [s_{1,f,t}, \ldots, s_{K,f,t}]^{\mathsf{T}} \in \mathbb{C}^K$, the multichannel observed signals $\boldsymbol{x}_{f,t}$ are modeled as

$$\boldsymbol{x}_{f,t} = \textstyle\sum_{k=1}^{K} \boldsymbol{a}_{k,f,t} s_{k,f,t} = A_{f,t}\boldsymbol{s}_{f,t}, \tag{1}$$

where $k = 1, \ldots, K$ is the source signal index, $f = 1, \ldots, F$ is the frequency bin index, and $t = 1, \ldots, T$ is the time frame index. Hereafter, ${}^{\mathsf{T}}, {}^{\mathsf{H}}$, and $\det$ denote vector/matrix transpose, Hermitian transpose and determinant, respectively. $\boldsymbol{a}_{k,f,t} \in \mathbb{C}^K$ is the steering vector that represents the

This work was supported by JST CREST Grant Number JPMJCR19A3 and JSPS Grant Number JP24K23896, Japan.

acoustic transfer characteristics of the kth source to each microphone, varying with time and frequency. The matrix $A_{f,t} = [\boldsymbol{a}_{1,f,t}, \dots, \boldsymbol{a}_{K,f,t}]$, composed of steering vectors, is known as the mixing matrix.

Online BSS aims to sequentially estimate the source signals from past and current observations as follows.

$$\boldsymbol{y}_{f,t} = W_{f,t}\boldsymbol{x}_{f,t}, \tag{2}$$

where $W_{f,t} = [\boldsymbol{w}_{1,f,t}, \dots, \boldsymbol{w}_{K,f,t}]^{\mathsf{H}} \in \mathbb{C}^{K \times K}$ is called the demixing matrix, which should be ideally the inverse of the mixing matrix.

III. Conventional Method

A. Online AuxIVA

Online AuxIVA [5] can be interpreted as estimating the demixing matrix $W_{f,t}$ at each time frame by minimizing the following objective function [14]:

$$J_t = \sum_{\tau=1}^{t}\sum_{k=1}^{K} \frac{\lambda^{t-\tau}}{\sum_{j=1}^{t}\lambda^{t-j}}\psi(r_{k,\tau}) - \sum_{f=1}^{F}\log|\det W_{f,t}|^2, \tag{3}$$

$$r_{k,\tau} = \sqrt{\textstyle\sum_{f=1}^{F}|\boldsymbol{w}_{k,f,\tau}^{\mathsf{H}}\boldsymbol{x}_{f,\tau}|^2}. \tag{4}$$

From (3)(4), $W_{f,t}$ is estimated using only the current and past observed signals $\{\boldsymbol{x}_{f,\tau}|\tau \le t\}$. The forgetting factor $\lambda \in [0,1)$ determines the weighting of past observations, with larger values preserving more past information. The contrast function $\psi(r_{k,\tau})$ is the negative log-likelihood function of the source signal and derived from the source model. In this study, assuming a time-varying complex Gaussian distribution for each source, the contrast function is given by $\psi(r_{k,\tau}) = r_{k,\tau}^2/2\sigma_{k,\tau}^2$, where $\sigma_{k,\tau}^2$ is the time-varying variance.

For efficiency, online AuxIVA estimates the demixing matrix $W_{f,t}$ by minimizing the following auxiliary function:

$$J_t^{+} = \sum_{f=1}^{F}\sum_{k=1}^{K}\boldsymbol{w}_{k,f,t}^{\mathsf{H}}V_{k,f,t}\boldsymbol{w}_{k,f,t} - \sum_{f=1}^{F}\log|\det W_{f,t}|^2, \tag{5}$$

$$V_{k,f,t} = \lambda V_{k,f,t-1} + (1-\lambda)\varphi(r_{k,t})\boldsymbol{x}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}, \tag{6}$$

$$r_{k,t} = \sqrt{\textstyle\sum_{f=1}^{F}|\hat{y}_{k,f,t}|^2}, \tag{7}$$

$$\hat{y}_{k,f,t} = \boldsymbol{w}_{k,f,t-1}^{\mathsf{H}}\boldsymbol{x}_{f,t}, \tag{8}$$

where $\varphi(r_{k,t})$ is defined as $\varphi(r_{k,t}) = \psi'(r_{k,t})/r_{k,t}$

using the first-order derivative $\psi'(r_{k,t})$ of the contrast function $\psi(r_{k,t})$. Also, $\hat{\boldsymbol{y}}_{f,t} = [\hat{y}_{1,f,t}, \dots, \hat{y}_{K,f,t}]^{\mathsf{T}} = W_{f,t-1}\boldsymbol{x}_{f,t}$ is the signal estimated using the demixing matrix of the previous time frame. The covariance matrix $V_{k,f,t}$ can be interpreted as a weighted average of the instantaneous covariance matrix $\boldsymbol{x}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}$ by $\varphi(r_{k,t}) = F/r_{k,t}^2$ in the case of the time-varying Gaussian source model. The weight $\varphi(r_{k,t})$ is designed to decrease as the source activity (7) increases. For example, $V_{i,f,t}$ is ideally like a covariance matrix consisting of sources other than the source $k = i$. The time-varying source variance is also updated by maximum-likelihood estimation by $\hat{\sigma}_{k,\tau}^2 = r_{k,\tau}^2/F$ in each iteration.

B. Demixing matrix update

To update the demixing matrix efficiently, the update rule, called the iterative projection (IP) [6], has been proposed. IP iteratively minimizes the objective function (5) for the demixing vector $\boldsymbol{w}_{k,f,t}(\forall k)$ using (9)–(11) at each time frame.

$$W_{f,t} \leftarrow W_{f,t-1}, \tag{9}$$

$$\boldsymbol{w}_{k,f,t} \leftarrow (W_{f,t}V_{k,f,t})^{-1}\boldsymbol{e}_k, \tag{10}$$

$$\boldsymbol{w}_{k,f,t} \leftarrow \frac{\boldsymbol{w}_{k,f,t}}{\sqrt{\boldsymbol{w}_{k,f,t}^{\mathsf{H}}V_{k,f,t}\boldsymbol{w}_{k,f,t}}}, \tag{11}$$

where $\boldsymbol{e}_k \in \mathbb{C}^K$ is a K-dimensional unit vector whose kth component should be 1 and all the other components should be 0. IP has no parameters such as step size, and each step can be calculated in a closed form.

IV. Proposed Method

In conventional online AuxIVA, the signal statistics, represented by $V_{k,f,t}$, is computed as a weighted covariance matrix of the observed signal $\boldsymbol{x}_{f,t}$ using a scalar forgetting factor. In contrast, this study reformulates the demixing matrix update as a multiplicative update algorithm, which enables source-dependent control of the forgetting factors by representing the signal statistics as a weighted covariance matrix of the separated signal $\hat{\boldsymbol{y}}_{f,t}$, rather than the observed signal $\boldsymbol{x}_{f,t}$.

A. Introducing multiplicative update into online AuxIVA

Previous studies [16] have formulated the demixing matrix update as a multiplicative update algorithm in batch AuxIVA. We extend this approach to the online AuxIVA. Here we consider updating the demixing matrix $W_{f,t}$ by multiplying the previous demixing matrix $W_{f,t-1}$ and the update matrix $Z_{f,t} = [\boldsymbol{z}_{1,f,t}, \dots, \boldsymbol{z}_{K,f,t}]^{\mathsf{H}} \in \mathbb{C}^{K \times K}$ as follows:

$$W_{f,t} = Z_{f,t}W_{f,t-1}. \tag{12}$$

Then, multiplying $W_{f,t-1}$ from the left and $W_{f,t-1}^{\mathsf{H}}$ from the right to (6), we obtain

$$\begin{aligned} W_{f,t-1}V_{k,f,t}W_{f,t-1}^{\mathsf{H}} &= \lambda W_{f,t-1}V_{k,f,t-1}W_{f,t-1}^{\mathsf{H}} \\ &\quad + (1-\lambda)\varphi(r_{k,t})W_{f,t-1}\boldsymbol{x}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}W_{f,t-1}^{\mathsf{H}}. \end{aligned} \tag{13}$$

Furthermore, by setting $\tilde{V}_{k,f,t} = W_{f,t-1}V_{k,f,t}W_{f,t-1}^{\mathsf{H}}$ and using (12), we obtain the update rule for the weighted covariance matrix in online AuxIVA with multiplicative update as follows.

$$\tilde{V}_{k,f,t} = \lambda Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}} + (1-\lambda)\varphi(r_{k,t})\hat{\boldsymbol{y}}_{f,t}\hat{\boldsymbol{y}}_{f,t}^{\mathsf{H}}. \tag{14}$$

Next, we derive the update rule for the update matrix $Z_{f,t}$. By multiplying (10) and (11) by $W_{f,t-1}^{-\mathsf{H}}$ from the left, we obtain

$$W_{f,t-1}^{-\mathsf{H}}\boldsymbol{w}_{k,f,t} \leftarrow (W_{f,t}V_{k,f,t}W_{f,t-1}^{\mathsf{H}})^{-1}\boldsymbol{e}_k, \tag{15}$$

$$W_{f,t-1}^{-\mathsf{H}}\boldsymbol{w}_{k,f,t} \leftarrow \frac{W_{f,t-1}^{-\mathsf{H}}\boldsymbol{w}_{k,f,t}}{\sqrt{\boldsymbol{w}_{k,f,t}^{\mathsf{H}}V_{k,f,t}\boldsymbol{w}_{k,f,t}}}. \tag{16}$$

From (12), since $\boldsymbol{w}_{k,f,t}^{\mathsf{H}} = \boldsymbol{z}_{k,f,t}^{\mathsf{H}} W_{f,t-1}$, it follows that

$$\boldsymbol{z}_{k,f,t} \leftarrow (Z_{f,t}\tilde{V}_{k,f,t})^{-1}\boldsymbol{e}_k, \tag{17}$$

$$\boldsymbol{z}_{k,f,t} \leftarrow \frac{\boldsymbol{z}_{k,f,t}}{\sqrt{\boldsymbol{z}_{k,f,t}^{\mathsf{H}}\tilde{V}_{k,f,t}\boldsymbol{z}_{k,f,t}}}, \tag{18}$$

where $Z_{f,t}$ is initialized by the identity matrix I at each f, t before applying (17) and (18).

B. Source-dependent forgetting factor

From (14), in online AuxIVA with multiplicative update, the signal statistics are expressed as the weighted covariance matrix of the separated signal. Since the source signals can be assumed to be statistically independent, the weighted covariance matrix of the separated signal is ideally expected to be a diagonal matrix. Based on this consideration, we introduce a diagonal forgetting factor matrix such as

$$\varLambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}, \tag{19}$$

where λ_k represents the forgetting factor for the kth source.

One way to introduce this source-dependent forgetting factor into the update rule (14) is as follows, ensuring that the quadratic form is preserved.

$$\begin{aligned}\tilde{V}_{k,f,t} &= \varLambda^{\frac{1}{2}} Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}}\varLambda^{\frac{1}{2}} \\ &\quad + \varphi(r_{k,t})(I-\varLambda)^{\frac{1}{2}}\hat{\boldsymbol{y}}_{f,t}\hat{\boldsymbol{y}}_{f,t}^{\mathsf{H}}(I-\varLambda)^{\frac{1}{2}},\end{aligned} \tag{20}$$

where the square root for a diagonal matrix is taken element-wise, applying the square root to each diagonal entry. Note that if the source-dependent forgetting factors are all the same ($\lambda_1 = \lambda_2 = \cdots = \lambda_K$), the update rule is equivalent to that of the conventional method.

Another way is to apply the source forgetting factor in an element-wise manner. We observe that the first term in (20) can also be expressed as

$$\begin{aligned}&\varLambda^{\frac{1}{2}} Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}}\varLambda^{\frac{1}{2}} = \\ &\qquad (\boldsymbol{p}\boldsymbol{p}^{\mathsf{T}}) \circ (Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}}),\end{aligned} \tag{21}$$

where $\boldsymbol{p} = \begin{bmatrix}\sqrt{\lambda_1} & \ldots & \sqrt{\lambda_K}\end{bmatrix}^{\mathsf{T}}$ and $\circ$ denotes the element-wise product of matrices. Thus, the update rule can be formulated using forgetting factors that sum to 1 per element, as follows.

$$\begin{aligned}\tilde{V}_{k,f,t} &= (\boldsymbol{p}\boldsymbol{p}^{\mathsf{T}}) \circ (Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}}) \\ &\quad + \varphi(r_{k,t})(\mathbf{1} - (\boldsymbol{p}\boldsymbol{p}^{\mathsf{T}})) \circ (\hat{\boldsymbol{y}}_{f,t}\hat{\boldsymbol{y}}_{f,t}^{\mathsf{H}}),\end{aligned} \tag{22}$$

where $\mathbf{1}$ is a matrix whose all elements are 1. However, preliminary experiments showed that using update rule (22), the separation performance often became unstable. This is because (22) is not expressed in a quadratic form, which may break the positive semidefiniteness of the covariance matrix. Therefore, in this study, we use (20) as the update rule. Algorithm 1 shows the process flow of the proposed online AuxIVA with multiplicative update, where N_{i} represents the number of iterations per time frame.

Algorithm 1 Online AuxIVA-IP with multiplicative update.

Initialize covariance matrices $V_{k,f,0}$ $(\forall k, f)$
Initialize demixing matrices $W_{f,0}$ $(\forall f)$
for $t = 1, \ldots, T$ **do**
  $W_{f,t} \leftarrow W_{f,t-1}$
  $\hat{\boldsymbol{y}}_{f,t} \leftarrow W_{f,t}\boldsymbol{x}_{f,t}$ $(\forall f)$
  for $i = 1, \ldots, N_{\mathrm{i}}$ **do**
    for $k = 1, \ldots, K$ **do**
      $r_{k,t} \leftarrow \sqrt{\sum_f |\boldsymbol{w}_{k,f,t}^{\mathsf{H}}\boldsymbol{x}_{f,t}|}$
      $\tilde{V}_{k,f,t} \leftarrow \varLambda^{\frac{1}{2}} Z_{f,t-1}\tilde{V}_{k,f,t-1}Z_{f,t-1}^{\mathsf{H}}\varLambda^{\frac{1}{2}} + \varphi(r_{k,t})(I-\varLambda)^{\frac{1}{2}}\hat{\boldsymbol{y}}_{f,t}\hat{\boldsymbol{y}}_{f,t}^{\mathsf{H}}(I-\varLambda)^{\frac{1}{2}}$ $(\forall f)$
    end for
    $Z_{f,t} \leftarrow I$ $(\forall f)$
    for $k = 1, \ldots, K$ **do**
      $\boldsymbol{z}_{k,f,t} \leftarrow (Z_{f,t}\tilde{V}_{k,f,t})^{-1}\boldsymbol{e}_k$ $(\forall f)$
      $\boldsymbol{z}_{k,f,t} \leftarrow \frac{\boldsymbol{z}_{k,f,t}}{\sqrt{\boldsymbol{z}_{k,f,t}^{\mathsf{H}}\tilde{V}_{k,f,t}\boldsymbol{z}_{k,f,t}}}$ $(\forall f)$
    end for
    $W_{f,t} \leftarrow Z_{f,t}W_{f,t-1}$ $(\forall f)$
  end for
end for

V. Experiment

A. Setup

We carried out numerical experiments to demonstrate the separation performance of the proposed method. In this study, the acoustic scene, such as source-microphone configuration, was fixed throughout the experiments to focus solely on evaluating the effectiveness of source-dependent forgetting. We obtained speech signals from ASJ Japanese Newspaper Article Sentences Read Speech Corpus (JNAS) [27] and concatenated them to make each signal $60\,\mathrm{s}$ long. All mixture signals were generated by convolving the impulse responses created by the gpuRIR simulation toolkit [17], based on the mirror image method, with the source signals. The reverberation time was approximately $300\,\mathrm{ms}$ and the sampling frequency was $16\,\mathrm{kHz}$. Fig. 1 shows the layout of the room. There were three sources, and only Source 1 moved from the start to the end point shown in Fig. 1 in $30\,\mathrm{s}$ to $2\,\mathrm{s}$. The microphone array was a three-channel circle arrangement with radius of $2\,\mathrm{cm}$ placed at the center of the room. Each microphone was omnidirectional. Under these conditions, we conducted 100 simulation experiments while randomly selecting speakers.

We considered the following two comparison methods. The first (Static) was the conventional online AuxIVA method using fixed forgetting factor. The second (Equal) was the proposed online AuxIVA with multiplicative update, in which the source-dependent forgetting factors in the forgetting factor matrix $\varLambda$ were all varied only in the moving segment as $\lambda_1 = \lambda_2 = \lambda_3$. In other words, it is equivalent to varying the scalar forgetting factor only for the moving segment in the conventional online AuxIVA. The proposed method (Proposed) also varied the forgetting factor corresponding to Source 1 while it was moving. In this experiment, the moving segment

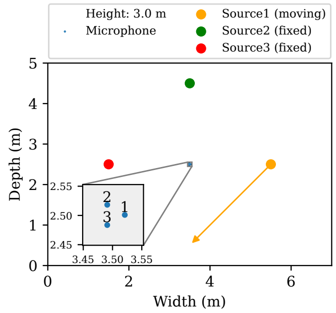

Fig. 1. Simulated room layout. Source 1 continuously moves in 30s to 2s. The microphone array and source 2 and 3 are fixed.

TABLE I
PARAMETERS IN THE EXPERIMENT.

Parameter	Value
Initial $\tilde{V}_{k,f,0}$	$I \times 10^{-3}$
Initial $W_{f,0}$	I
Sampling frequency	16 kHz
STFT window length	4096 (256 ms)
STFT window shift	2048 (128 ms)

TABLE II
MEAN SI-SDRI [dB] IN THE SIMULATION EXPERIMENTS.

	Static	Equal			Proposed		
$\tilde{\lambda}$		0.90	0.80	0.50	0.90	0.80	0.50
Source 1	2.79	3.99	4.51	5.27	4.14	6.32	**6.53**
Source 2	4.12	5.25	5.20	5.67	5.11	6.96	**7.38**
Source 3	5.37	6.66	6.36	5.57	6.29	**6.77**	5.97

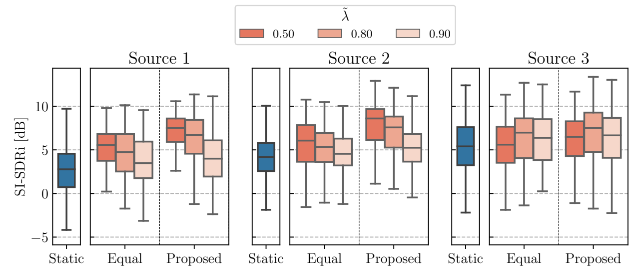

Fig. 2. Boxplot of SI-SDRi [dB] for each method.

and the index ℓ corresponding to Source 1 were given as oracle. The index ℓ were obtained by performing online AuxIVA with fixed forgetting factor, and then mapping the separated sources to each source using the Signal-to-Interference Ratio (SIR). The forgetting factors before Source 1 moved were $\lambda = 0.98$ for Static and Equal, and $\lambda_1 = \lambda_2 = \lambda_3 = 0.98$ for Proposed. In other words, before Source 1 movement (before 30 s), Static, Equal, and Proposed are all equivalent to each other. This setting was determined based on preliminary experiments, which showed that $\lambda = 0.98$ provided the best performance for stationary sources. In the moving segment, the forgetting factor was set to $\lambda = 0.90, 0.80, 0.50$ in Equal and $\lambda_\ell = 0.90, 0.80, 0.50$ in Proposed, where λ_ℓ is the forgetting factor corresponding to Source 1. Henceforth, we denote the forgetting factor in moving segment as $\tilde{\lambda}$. Other parameters are shown in Table I.

We evaluated the separation performance in terms of the scale-invariant signal-to-distortion ratio improvement (SI-SDRi) [18]. The scale of the separated signal $y_{k,f,t}$ was restored by projection back [19] onto the first microphone.

B. Result

Table II shows the average SI-SDRi after Source 1 moved (after 30 s) in the simulation experiment. This reveals that the proposed method with the source-dependent forgetting factor control improves the overall separation performance compared to the conventional method. When the forgetting factor of the moving segment $\tilde{\lambda} = 0.50$, the proposed method showed the best separation performance among all the methods.

Boxplot of SI-SDRi after the moving segment are shown in Fig. 2. From this, the separation performance of Source 1 corresponding to the moving source is greatly improved, especially when $\tilde{\lambda} = 0.50$. The covariance matrix $\tilde{V}_{\ell,f,t}$ corresponding to Source 1 is weighted by (7) and should therefore only contain statistics for Sources 2 and 3, the fixed sources, ideally. Thus, the covariance matrix $\tilde{V}_{\ell,f,t}$ does not need to be forgotten. However, the conventional method adjusts the signal statistics of all sources with a scalar weight, which may result in forgetting the signal statistics that does not need to be forgotten, and may degrade the separation performance. On the other hand, by controlling only the forgetting factor λ_ℓ corresponding to Source 1, the proposed method can forget only the signals statistics that should be forgotten, which is assumed to have contributed to the separation performance improvement.

In order to compare Equal and Proposed, Fig. 3 shows the scatter plot of SI-SDRi after the Source 1 movement when $\tilde{\lambda} = 0.50$. The majority of the points lie above the diagonal, which indicates that Proposed method outperforms Equal. In detail, the separation performance improved by 75/100 for Source 1, 73/100 for Source 2, and 65/100 for Source 3.

Fig. 4 shows the temporal variation of SI-SDRi for each source in one simulation. Before Source 1 moved, Proposed is equivalent to Static when the source-dependent forgetting factors λ_k $(\forall k)$ are all the same value. On the other hand, after Source 1 have moved, Proposed shows better separation performance than others, suggesting improved performance in dynamic environments such as when sound sources move.

These results confirm the effectiveness of the proposed source-dependent forgetting factor control. Although the source index and moving segment were given as oracle in this experiments, we also evaluated the method's robustness. Even with a 1 s delay in the moving segment annotation, the proposed method outperformed conventional methods, indicating a certain degree of robustness to timing errors.

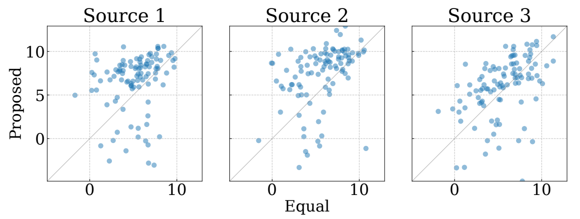

Fig. 3. Scatterplot of SI-SDRi [dB] for Equal and Proposed when $\tilde{\lambda} = 0.50$.

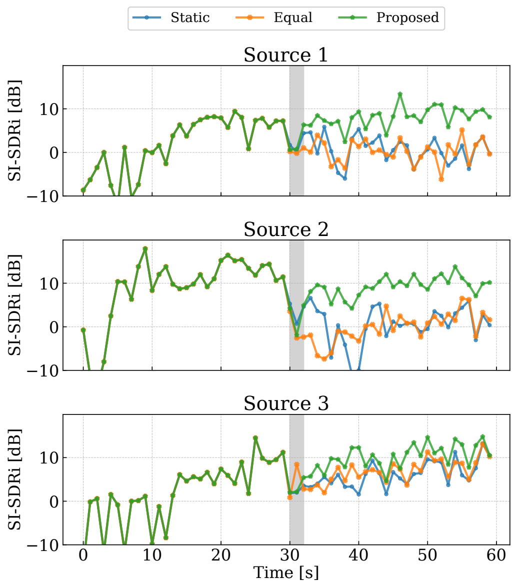

Fig. 4. Temporal variation of SI-SDRi [dB] for each method when $\tilde{\lambda} = 0.50$. The gray shading indicates the moving segment of Source 1.

VI. Conclusion

In this paper, we introduced a multiplicative update formulation for online AuxIVA, which enables source-dependent forgetting factor control. The proposed method allows source-wise adjustment of signal statistics, offering more flexibility than conventional methods. Simulation experiments demonstrated that the proposed method significantly improves separation performance, particularly in dynamic environments such as when sound sources move. In this experiment, the moving segment and the index ℓ corresponding to the moving source were given as oracle, but in the future, we will explore adaptive estimation of the forgetting factor to dynamically adjust to source movement without prior knowledge.

References

[1] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2006.

[2] A. Hiroe, “Solution of permutation problem in frequency domain ICA, using multivariate probability density functions,” in *Proceedings of International Conference on Independent Componet Analysis and Blind Signal Separation (ICA)*, Mar. 2006, pp. 601–608.

[3] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, “A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments,” *Signal Processing*, vol. 86, no. 6, pp. 1260–1277, 2006.

[4] T. Kim, “Real-time independent vector analysis for convolutive blind source separation,” *IEEE Trans. Circuits Syst. I*, vol. 57, no. 7, pp. 1431–1438, 2010.

[5] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, “An auxiliary-function approach to online independent vector analysis for real-time blind source separation,” in *Proceedings of Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, May 2014, pp. 107–111.

[6] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2011, pp. 189–192.

[7] M. Sunohara, C. Haruta, and N. Ono, “Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 216–220.

[8] Y. Kuriki, T. Nakashima, K. Yamaoka, N. Ueno, Y. Wakabayashi, N. Ono, and R. Sato, “Efficient low-latency convolution with uniform filter partition and its evaluation on real-time blind source separation,” in *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2022, pp. 765–769.

[9] K. Mo, X. Wang, Y. Yang, S. Makino, and J. Chen, “Low algorithmic delay implementation of convolutional beamformer for online joint source separation and dereverberation,” in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2024, pp. 912–916.

[10] T. Ueda, T. Nakatani, R. Ikeshita, K. Kinoshita, S. Araki, and S. Makino, “Low latency online blind source separation based on joint optimization with blind dereverberation,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 506–510.

[11] ——, “Low latency online source separation and noise reduction based on joint optimization with dereverberation,” in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1000–1004.

[12] S. Yuan, T. Ueda, and S. Makino, “Real-time moving blind source extraction based on constant separating vector and auxiliary function technique,” *Journal of Signal Processing*, vol. 27, no. 4, pp. 81–85, 2023.

[13] T. Nakashima and N. Ono, “Inverse-free online blind source separation based on independent vector analysis with flexible iterative source steering,” in *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2022, pp. 750–754.

[14] 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 *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5.

[15] T. Nakashima, Y. Wakabayashi, and N. Ono, “Self-rotation-robust online independent vector analysis with sound field interpolation on circular microphone array,” *APSIPA Trans. Signal and Information Processing*, 2023.

[16] R. Ikeshita and N. Tomohiro, “ISS2: an extension of iterative source steering algorithm for majorization-minimization-based independent vector analysis,” in *Proceedings of European Signal Processing Conference (EUSIPCO)*, Aug. 2022, pp. 65–69.

[17] D. Diaz-Guerra, A. Miguel, and J. R. Beltran, “gpuRIR: A python library for room impulse response simulation with gpu acceleration,” *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5653–5671, Oct. 2020.

[18] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, “SDR — half-baked or well done?” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 626–630.

[19] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, Oct. 2001.