2023 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (RISP NCSP 2023), Feb. 28 - Mar. 3, 2023

Investigation of Partial Update and Source Localization in Geometrically Constrained Independent Vector Analysis with Auxiliary Function Approach for Moving Source

Kana Goto[1], Tetsuya Ueda[2], Li Li[3], Takeshi Yamada[1], and Shoji Makino[2,1]

[1]University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
E-mail: goto.kana.ry@alumni.tsukuba.ac.jp, takeshi@cs.tsukuba.ac.jp

[2]Waseda University
2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan
E-mail: t.ueda@akane.waseda.jp, s.makino@waseda.jp

[3] NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
3-1 Morinosato-Wakamiya, Kanagawa 243-0198, Japan
E-mail: lili-0805@ieee.org

Abstract

In this paper, we apply an efficient partial update to online geometrically constrained independent vector analysis with auxiliary function approach and iterative source steering (ISS), "oGC-AuxIVA-ISS" for short. The algorithm fully exploits the advantages of the auxiliary function approach, i.e., fast convergence and no stepsize tuning, and ISS, i.e., low computational complexity and numerical stability, making it highly suitable for practical use. Moreover, ISS theoretically provides a more efficient way to deal with moving sources by partially updating the demixing filters corresponding to the sources, which can further improve computational efficiency. We conduct moving source separation experiments to confirm its effect on the online algorithm. Moreover, we adopt a method based on delay-and-sum (DS) to estimate the directions of arrival (DOAs) of sources. Experimental results showed that oGC-AuxIVA-ISS with partial updates outperformed the conventional method without ISS with partial updates, and the DOA estimation based on DS could further reduce runtime.

1. Introduction

The presence of background noise and directional interferences can severely degrade the quality of recorded speech and subsequently decrease the performance of speech-targeted applications. This raises the need for techniques to separate the target speech from other sounds. Especially for face-to-face applications, e.g., hearing-aid devices and teleconference systems, it is necessary to develop real-time source separation systems, where processes for signals in the current frame have to be finished before the next frame arrives.

Geometrically constrained blind source separation (GC-BSS) [1–6] allows us to separate signals and select the desired source by combining the optimization problem of BSS [7–11] with spatial constraints. Two online extensions of GC-BSS, online geometrically constrained independent vector analysis with auxiliary function approach and vectorwise coordinate descent (oGC-AuxIVA-VCD) [2] and online GC-AuxIVA with iterative source steering (oGC-AuxIVA-ISS) [6], fully exploit the advantages of the auxiliary function approach [12, 13], i.e., fast convergence and no stepsize tuning, making them more suitable for practical applications. oGC-AuxIVA-VCD updates the demixing filters by solving hybrid exact-approximate diagonalization problems. oGC-AuxIVA-ISS, on the other hand, uses ISS [11] to update the demixing matrix with a sequence of rank-1 operations, leading to an inverse-free algorithm. Furthermore, thanks to the rank-1 updates, ISS provides a more efficient way to deal with moving sources by partially updating the demixing filters corresponding to them [14]. In this paper, to improve computational efficiency of oGC-AuxIVA-ISS, we first refer to the update as "partial update" and examine the effectiveness of it in oGC-AuxIVA-ISS for moving sound source separation.

Secondly, we investigate the performance impact using estimated DOAs of the moving sound source. Knowing the exact directions of arrival (DOA) of sources is a strong constraint in practical applications for GC-BSS. So far, several studies have used DOA estimation for GC-BSS [2, 6, 15] to address this limitation. In this paper, we investigate and compare DOA estimation method based on the delay-and-sum (DS) beamformer [16], multiple signal classification (MUSIC) [17], and generalized cross correlation with phase transform (GCC-PHAT) [18] as promising methods to improve the DOA estimation efficiency. We also introduce a scale constraints [5, 19] to both methods to further improve numerical stability.

2. Conventional oGC-AuxIVA-VCD and oGC-AuxIVA-ISS

2.1 AuxIVA with geometric constraints and scale constraints

Let us consider a determined situation where J sources are observed by I microphones. Here, $I = J$. Let x_{ifn} and y_{jfn} denote the short-time Fourier transform (STFT) coefficients of the signal observed at the ith microphone and that output at the jth channel, respectively. Here, $f = 1, \ldots, F$ and $n = 1, \ldots, N$ are the indices of the frequency and time frame, respectively. We denote the frequency-wise vector representation of the observed and the estimated signals by

$$\boldsymbol{x}_{fn} = [x_{1fn}, \ldots, x_{Ifn}]^{\mathsf{T}} \in \mathbb{C}^{I}, \tag{1}$$

$$\boldsymbol{y}_{fn} = [y_{1fn}, \ldots, y_{Jfn}]^{\mathsf{T}} \in \mathbb{C}^{J}, \tag{2}$$

where $(\cdot)^{\mathsf{T}}$ denotes the transpose. When considering a time-variant instantaneous mixture model, the relationship between the observed and estimated signals can be expressed as

$$\boldsymbol{y}_{fn} = \boldsymbol{W}_{fn}\boldsymbol{x}_{fn}. \tag{3}$$

Here, $\boldsymbol{W}_{fn} = [\boldsymbol{w}_{1fn}, \ldots, \boldsymbol{w}_{Jfn}]^{\mathsf{H}}$ is an $I \times I$ demixing matrix containing demixing filters $\boldsymbol{w}_{jfn} = [w_{1jfn}, \ldots, w_{Ijfn}]^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose.

IVA [8, 9] assumes that each frame of source follows a multivariate distribution, and thus dependencies over frequency components can be exploited to solve frequency-domain permutation problem simultaneously with frequency-wise source separation. The demixing matrices $\boldsymbol{W}_{fn}$ are estimated by minimizing the following negative log-likelihood function:

$$\mathcal{L}_{\text{IVA}} = \sum_{j=1}^{J} \mathbb{E}[G(\boldsymbol{y}_{jn})] - \sum_{f=1}^{F} \log|\det \boldsymbol{W}_{fn}|, \tag{4}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator over frames and $\boldsymbol{y}_{jn} = [y_{j1n}, \ldots, y_{jFn}]^{\mathsf{T}} \in \mathbb{C}^F$ is the source-wise vector representation. Here, $G(\boldsymbol{y}_{jn})$ is the contrast function having the relationship $G(\boldsymbol{y}_{jn}) = -\log p(\boldsymbol{y}_{jn})$. One typical choice of the contrast function is to use a spherical contract function [8–10], which is expressed as

$$G(\boldsymbol{y}_{jn}) = G_R(r_{jn}), \tag{5}$$

$$r_{jn} = ||\boldsymbol{y}_{jn}||_2 = \sqrt{\sum_f |y_{jfn}|^2} = \sqrt{\sum_f |\boldsymbol{w}_{jfn}^{\mathsf{H}} \boldsymbol{x}_{fn}|^2}. \tag{6}$$

Here, $G_R(r)$ is a function of a real-valued scalar variable r and $||\cdot||_2$ denotes the L_2 norm. By adopting the auxiliary function approach [10, 20], an upper bound is optimized instead of the original objective function, which is expressed as

$$\begin{aligned}\mathcal{L}_{\text{IVA}} &\leq \mathcal{L}_{\text{AuxIVA}} \\ &= \frac{1}{2}\sum_{f=1}^{F}\sum_{j=1}^{J} \boldsymbol{w}_{jfn}^{\mathsf{H}} \boldsymbol{\Sigma}_{jfn} \boldsymbol{w}_{jfn} - \sum_{f=1}^{F} \log|\det \boldsymbol{W}_{fn}|. \end{aligned} \tag{7}$$

Here, $\boldsymbol{\Sigma}_{jfn}$ is a weighted spatial covariance matrix and is autoregressively calculated at each frame n using the previously calculated one [21] since only the observed signals up to the present time are available in the case of online processing:

$$\boldsymbol{\Sigma}_{jfn} = \alpha \boldsymbol{\Sigma}_{jf(n-1)} + (1-\alpha)\varphi(r_{jn}) \boldsymbol{x}_{fn} \boldsymbol{x}_{fn}^{\mathsf{H}}. \tag{8}$$

Here, $\varphi(r_{jn}) = G_R^{'}(r_{jn})/r_{jn}$, where $(\cdot)'$ denotes the derivative operator, and $0 \leq \alpha < 1$ denotes a forgetting factor.

Now, let us consider geometric constraints (GC) [22] that restrict the far-field response of filters estimated by IVA in a set of directions Θ_{jn}, which is described as

$$\mathcal{L}_{\text{GC}} = \sum_{f=1}^{F}\sum_{j=1}^{J}\sum_{\theta\in\Theta_{jn}} \lambda^{\text{GC}} |\boldsymbol{w}_{jfn}^{\mathsf{H}} \boldsymbol{d}_{f\theta} - c_{j\theta}|^2. \tag{9}$$

Here, Θ_{jn} denotes a set including all directions to be considered, $\boldsymbol{d}_{f\theta}$ is the steering vector pointing to the direction θ, $c_{j\theta}$ is a nonnegative value set for all frequency bins as constraints, and $\lambda^{\text{GC}} \geq 0$ is a parameter that weighs the importance of the constraint. Note that (9) with $c_{j\theta} = 1$ forces the spatial filter to form a conventional delay-and-sum (DS) beamformer steering in the direction θ to preserve the target source, whereas a small value of $c_{j\theta}$ essentially creates a spatial null towards the direction θ so that multiple constraints of spatial nulls towards the directions of all interferences can be used to suppress all interferences and preserve the target. Note that no auxiliary function is required since these geometric constraints are linear and can be easily optimized.

In addition, the source separation method based on source independence may cause numerical instability due to the scale ambiguity problem. Therefore, in this paper, we consider scale constraints (SC) [5, 19] expressed as

$$\mathcal{L}_{\text{SC}} = \lambda^{\text{SC}} \sum_{f=1}^{F}\sum_{j=1}^{J} \boldsymbol{w}_{jfn}^{\mathsf{H}} \boldsymbol{w}_{jfn}, \tag{10}$$

where λ^{SC} is a parameter that weighs the importance of the constraint. This is expected to induce the demixing filter to be scaled down, thus stabilizing the numerical computation.

Therefore, the auxiliary function for GC-AuxIVA, a.k.a., the objective function to be minimized, is given as

$$\mathcal{L} = \mathcal{L}_{\text{AuxIVA}} + \mathcal{L}_{\text{GC}} + \mathcal{L}_{\text{SC}}. \tag{11}$$

Hereafter, to derive the update rules easily, the index of f is omitted.

2.2 Online GC-AuxIVA-VCD [2]

The update rule for the weighted spatial covariance matrix $\boldsymbol{\Sigma}_{jn}$ is obtained straightforwardly by substituting (6) into (8) and those for $\boldsymbol{W}_n$ are derived by embracing the idea adopted in VCD [23], which are summarized as follows:

$$\boldsymbol{u}_{jn} = \boldsymbol{D}_{jn}^{-1} \boldsymbol{W}_n^{-1} \boldsymbol{e}_j, \tag{12}$$

$$\hat{\boldsymbol{u}}_{jn} = \lambda^{\text{GC}} \boldsymbol{D}_{jn}^{-1} \sum_{\theta\in\Theta_{jn}} c_{j\theta} \boldsymbol{d}_\theta, \tag{13}$$

$$h_{jn} = \boldsymbol{u}_{jn}^{\mathsf{H}} \boldsymbol{D}_{jn} \boldsymbol{u}_{jn}, \tag{14}$$

$$\hat{h}_{jn} = \boldsymbol{u}_{jn}^{\mathsf{H}} \boldsymbol{D}_{jn} \hat{\boldsymbol{u}}_{jn}, \tag{15}$$

$$\boldsymbol{w}_{jn} = \begin{cases} \frac{1}{\sqrt{h_{jn}}} \boldsymbol{u}_{jn} + \hat{\boldsymbol{u}}_{jn} \ \ (\text{if } \hat{h}_{jn} = 0), \\ \frac{\hat{h}_{jn}}{2h_{jn}} \Big[-1 + \sqrt{1 + \frac{4h_{jn}}{|\hat{h}_{jn}|^2}} \Big] \boldsymbol{u}_{jn} + \hat{\boldsymbol{u}}_{jn} \ \ (\text{o.w.}). \end{cases} \tag{16}$$

Here, $\boldsymbol{D}_{jn} = \boldsymbol{\Sigma}_{jn} + \lambda^{\text{SC}} \boldsymbol{I} + \lambda^{\text{GC}} \sum_{\theta\in\Theta_{jn}} \boldsymbol{d}_\theta \boldsymbol{d}_\theta^{\mathsf{H}}$, $\boldsymbol{I}$ is the $I \times I$ identity matrix, and $\boldsymbol{e}_j$ is the jth column of $\boldsymbol{I}$.

2.3 Online GC-AuxIVA-ISS [6]

As (12) shows, either offline or online GC-AuxIVA-VCD requires the matrix inversion for each frequency, source, and iteration, which is typically considered to be computationally expensive and numerically unstable, and therefore should be avoided in practice. To address this, an alternative algorithm for online GC-AuxIVA by replacing VCD with the recently proposed ISS is proposed.

Instead of updating each row of the demixing matrix $\boldsymbol{W}_n$ alternately, ISS performs a rank-1 update for the whole demixing matrix as

$$\boldsymbol{W}_n \leftarrow \boldsymbol{W}_n - \boldsymbol{v}_{jn} \boldsymbol{w}_{jn}^{\mathsf{H}} \tag{17}$$

for $j = 1, \ldots, I$. Here, $\boldsymbol{v}_{jn} = [v_{1jn}, \ldots, v_{Ijn}]^{\mathsf{T}} \in \mathbb{C}^I$ is a vector to be estimated instead of the demixing matrix.

Substituting (17) into the objective function of online GC-AuxIVA-VCD, i.e., (11) with time-varying demixing filters $\boldsymbol{w}_{jn}$ and looking directions Θ_{jn}, we have the new objective function to be minimized. By solving partial derivative of it w.r.t. v_{ijn}^*, we obtain following update rules:

$$v_{ijn} = \frac{\boldsymbol{w}_{in}^{\mathsf{H}} (\boldsymbol{\Sigma}_{in} + 2\lambda^{\text{SC}} \boldsymbol{I}) \boldsymbol{w}_{jn} + 2\lambda^{\text{GC}} \sum_{\theta\in\Theta_{jn}} g_{j\theta}^* (g_{i\theta} - c_{i\theta})}{\boldsymbol{w}_{jn}^{\mathsf{H}} (\boldsymbol{\Sigma}_{in} + 2\lambda^{\text{SC}} \boldsymbol{I}) \boldsymbol{w}_{jn} + 2\lambda^{\text{GC}} \sum_{\theta\in\Theta_{jn}} |g_{j\theta}|^2} \quad (\forall i \neq j) \tag{18}$$

$$v_{jjn} = \begin{cases} 1 - p_{jn}^{-1/2} \ (\text{if } q_{jn} = 0), \\ 1 - q_{jn}^* \frac{|q_{jn}| + \sqrt{|q_{jn}|^2 + p_{jn}}}{p_{jn}|q_{jn}|} \ (\text{o.w.}), \end{cases} \tag{19}$$

where,

$$g_{j\theta} = \boldsymbol{w}_{jn}^{\mathsf{H}} \boldsymbol{d}_\theta, \tag{20}$$

$$p_{jn} = \boldsymbol{w}_{jn}^{\mathsf{H}} (\boldsymbol{\Sigma}_{jn} + 2\lambda^{\text{SC}} \boldsymbol{I}) \boldsymbol{w}_{jn} + 2\lambda^{\text{GC}} \sum_{\theta\in\Theta_{jn}} |g_{j\theta}|^2, \tag{21}$$

$$q_{jn} = \lambda^{\text{GC}} \sum_{\theta\in\Theta_{jn}} c_{j\theta} g_{j\theta}. \tag{22}$$

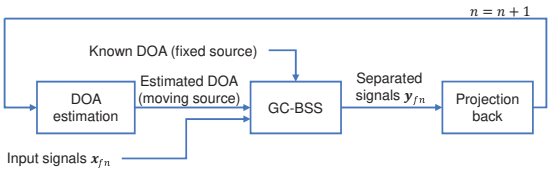
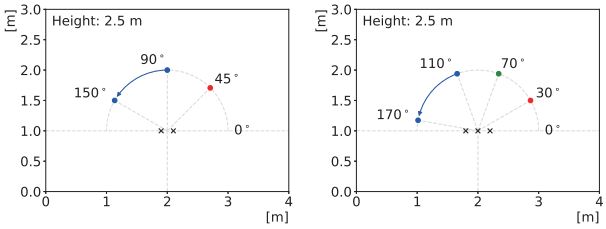


Figure 1: Update flow of the system combining GC-BSS and DOA estimation



(a) 2-speaker case (b) 3-speaker case

Figure 2: Layout of sound sources and microphones, where "○" and "×" denote source and microphone positions, respectively.

3. Partial update and DOA estimation for oGC-AuxIVA-ISS

3.1 Partial update

This paper investigates the performance of partial update [14] when applied to online GC-AuxIVA-ISS to improve computational efficiency. Considering the situation where only the jth source is moving, namely, the jth steering vector is time-variant and the other steering vectors are time-invariant, which do not need to update after convergence. Thanks to the fact that the 1-rank update of ISS is equivalent to updating the steering vector corresponding to each source, which is expressed as

$$\boldsymbol{a}_{jn} \leftarrow \frac{1}{1 - v_{jjn}} \left(\boldsymbol{a}_{jn} + \sum_{i \neq j} v_{ijn} \boldsymbol{a}_{in} \right), \tag{23}$$

we are allowed to perform more efficient update in the above moving source situation by only update v_{ijn} $(i = 1, \dots I)$ corresponding to the moving source while keeping others fixed after convergence.

3.2 DOA estimation

In addition, we consider a system that uses the spatial information contained in the demixing filter obtained by updating the GC-BSS for DOA estimation. The system is shown in Fig. 1, where the process in each frame can be summarized as follows: 1) estimate the DOA of the moving source using the source image calculated at the last frame; 2) input the known DOA and the estimated DOA to GC-BSS for source separation; 3) generate source images using projection back technique for the DOA estimation in the next frame. By iteratively doing these processes in each frame, the separated signal can be sequentially obtained while controlling the output order.

For DOA estimation methods, this paper compares DS-based DOA estimation [16], multiple signal classification (MUSIC) [17], and generalized cross correlation with phase transform (GCC-PHAT) [18].

DS is a technique to enhance signals arriving from a certain direction by aligning the phase of the wave using the time difference of arrival (TDOA) between the microphones. It calculates the angle θ_{jn} that maximizes the cross-correlation between the steering vector $\boldsymbol{a}_{jn}$ obtained from the estimated mixing matrix $\boldsymbol{A}_n = \boldsymbol{W}_n^{-1}$ and the steering vector $\boldsymbol{d}_\theta$ based on the assumption of the arrival of plane waves:

$$\theta_{jn} = \frac{2}{F} \sum_{f=1}^{F/2} \operatorname*{argmax}_{\theta} |\boldsymbol{a}_{jn}^{\mathsf{H}} \boldsymbol{d}_\theta|^2. \tag{24}$$

In each frame n, we calculate all angles θ_{jn} $(j = 1, \dots, J)$ and take the farthest angle from the known angle as the DOA of the moving source since we only consider one unknown DOA.

MUSIC is a subspace-based method that decomposes the covariance matrix of the observed multichannel signals to obtain subspaces of signals and noise that are orthogonal to each other. The operation of MUSIC in this system is described in [6]. We used [24] for implementing the MUSIC method.

GCC-PHAT estimates the TDOA of the incident wave between microphone pairs using a generalized cross-correlation function. The system estimates TDOA τ_{jn} as follows:

$$\tau_{jn} = \operatorname*{argmax}_{\tau} \sum_{f=1}^{F/2} \frac{\tilde{y}_{j1fn} \tilde{y}_{j2fn}^*}{|\tilde{y}_{j1fn} \tilde{y}_{j2fn}^*|} \exp(2\sqrt{-1}\pi f \tau), \tag{25}$$

where y_{jkfn} is the source image corresponding to the kth microphone of the separated signal y_{jfn} obtained by projection back. Since τ_{jn} can be calculated based on the microphone arrangement, GCC-PHAT can calculate the target angle.

4. Experimental evaluations

We conducted source separation experiments containing a moving source to evaluate the effectiveness of partial updates.

We used speech signals of 6 speakers (3 males and 3 females) extracted from the ATR Japanese Speech Database [25]. By randomly selecting 2 or 3 speakers from the database, we generated 20 mixture signals with length of 60 seconds for each. We used the `signal generator`[1] to simulate room impulse responses (RIRs), and the layout of sound sources and microphones is shown in Fig. 2. In both the 2-speaker and 3-speaker cases, the moving speaker was fixed for the first 20 seconds, moved on an arc for the next 20 seconds, and finally fixed for the last 20 seconds. The number of microphones was set equal to the number of sound sources. The interval of microphones was set at 2 cm. The reverberation time (RT_{60}) was set at 200 ms. All the speech signals were sampled at 16 kHz. The STFT was computed using a Hanning window, whose length and shift were set at 1024 samples (64 ms) and 512 samples (32 ms), respectively. We initialized $\boldsymbol{\Sigma}_{jf0}$ and $\boldsymbol{W}_{f0}$ as identity matrices and set the forgetting parameter α at 0.99 for each method. In the DOA estimation of the first frame, we used the source image of the observed signals instead of the separated signals. We adopted the null constraint, where Θ_{jn} is the set of DOAs of the signal to be suppressed by the jth demixing filter. We investigated several values of λ^{GC} and λ^{SC}, and chose the optimal one experimentally.

To confirm the effectiveness of the partial update, we considered two update rules, "one" and "all". In "all", the update rules, (12)–(16) in oGC-AuxIVA-VCD or (18) and (19) in oGC-AuxIVA-ISS, were applied for all $j = 1, \dots, J$ throughout the observation. In "one", the update rules were applied for all $j = 1, \dots, J$ in the first 20 seconds and then applied for one specific j corresponding to the moving source in the remaining 40 seconds. We assumed that the DOAs of the fixed sources are known in all frames, and we estimated

[1] https://www.audiolabs-erlangen.de/fau/professor/habets/software/signalgenerator

Table 1: Average runtime [s] for 60 second signals.

DOA estimation	oGC-AuxIVA-VCD "all"	oGC-AuxIVA-VCD "one"	oGC-AuxIVA-ISS "all"	oGC-AuxIVA-ISS "one"
2-speaker case				
DS	32.66	18.76	13.36	12.01
MUSIC	76.43	67.96	57.19	55.54
GCC-PHAT	37.55	35.23	26.74	25.74
3-speaker case				
DS	64.82	45.33	31.03	26.41
MUSIC	154.05	134.92	120.45	115.72
GCC-PHAT	69.51	51.99	37.21	32.40

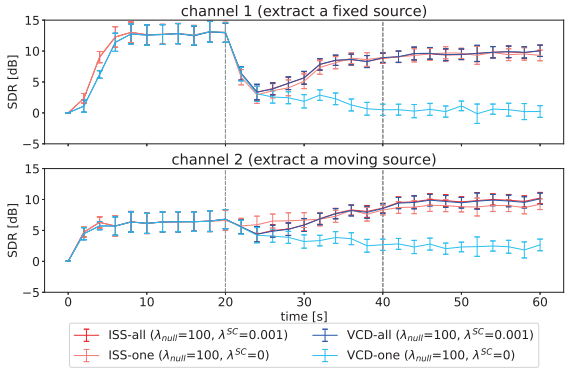
Table 2: Average of SDR [dB] every 2 seconds for each method.

DOA estimation	oGC-AuxIVA-VCD "all"	oGC-AuxIVA-VCD "one"	oGC-AuxIVA-ISS "all"	oGC-AuxIVA-ISS "one"
2-speaker case				
DS	8.29	4.30	8.35	8.20
MUSIC	8.37	4.37	8.33	8.26
GCC-PHAT	8.47	4.60	8.47	8.28
3-speaker case				
DS	4.57	0.93	4.55	3.47
MUSIC	4.46	0.47	4.02	2.61
GCC-PHAT	4.53	0.56	4.56	3.51

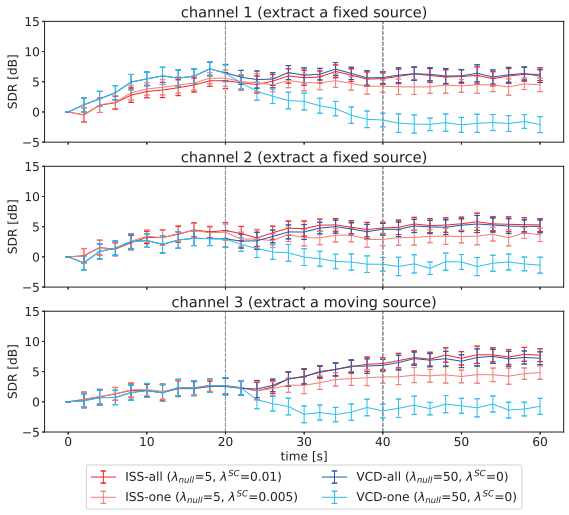
the DOA of the moving source. The enhancement performance was evaluated using the source-to-distortions ratio (SDR) [26].

First, regarding the update rules, "one" had shorter runtime than "all" as shown in Table 1. Next, in terms of the DOA estimation methods, DS was the fastest for the computation, followed by GCC-PHAT, and MUSIC. In addition, DS and GCC-PHAT showed almost equal or higher SDR than MUSIC, as shown in Table 2. Therefore, DS was the best DOA estimation method in terms of separation performance and computational efficiency in this experiment.

Next, Fig. 3 shows the SDR every 2 seconds with DS. In Fig. 3a, the 2-speaker case, oGC-AuxIVA-ISS-all and oGC-AuxIVA-one showed almost the same scores as oGC-AuxIVA-VCD-all. On the other hand, as expected from the update rules, the score of oGC-AuxIVA-VCD-one decreased significantly from around 20 seconds after the start of partial update. In Fig. 3b, the 3-speaker case, oGC-AuxIVA-ISS-all and oGC-AuxIVA-VCD-all showed similar SDR scores, while oGC-AuxIVA-ISS-one showed a slightly lower score. The score of oGC-AuxIVA-VCD-one decreased significantly from around 20 seconds after the start of partial updating, as was the 2-speaker case.



(a) 2-speaker case



(b) 3-speaker case

Figure 3: Average SDR [dB] for each channel with DS in every 2 seconds.

5. Conclusions

In this paper, we applied an efficient partial update to oGC-AuxIVA-ISS and evaluated its effectiveness for moving source separation when using DOA estimation. Experimental results showed that the separation performance of oGC-AuxIVA-ISS with partial updates was higher than that of oGC-AuxIVA-VCD with partial updates. Furthermore, it was shown that DOA estimation based on DS could further reduce runtime.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19H04131.

References

[1] A. H. Khan *et al.*, in *Proc. LVA/ICA*, 2015, pp. 396–403.
[2] L. Li *et al.*, in *Proc. Interspeech*, 2020, pp. 61–65.
[3] L. C. Parra *et al.*, *IEEE Trans. SAP*, vol. 10, no. 6, pp. 352–362, 2002.
[4] H. Saruwatari *et al.*, *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.
[5] A. Brendel *et al.*, *IEEE Trans. Signal Processing*, vol. 68, pp. 3545–3558, 2020.
[6] K. Goto *et al.*, in *Proc. APSIPA*, 2022, pp. 755–760.
[7] A.Hyvärinen *et al.*, *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
[8] T. Kim *et al.*, in *Proc. ICA*, 2006, pp. 165–172.
[9] A. Hiroe, in *Proc. ICA*, 2006, pp. 601–608.
[10] N. Ono, in *Proc. WASPAA*, 2011, pp. 189–192.
[11] R. Scheibler *et al.*, in *Proc. ICASSP*, 2020, pp. 236–240.
[12] N. Ono *et al.*, in *Proc. LVA/ICA*, 2010, pp. 165–172.
[13] N. Ono, in *Proc. IWAENC*, 2012, pp. 1–4.
[14] T. Nakashima *et al.*, in *Proc. APSIPA*, 2022, pp. 185–188.
[15] A. Lombard *et al.*, in *Proc. ICASSP*, 2009, pp. 233–236.
[16] K. Mogi *et al.*, in *Proc. Autumn Meeting of Acoustical Society of Japan*, 2021, pp. 153–154.
[17] R. Schmidt, *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
[18] C. Knapp *et al.*, *IEEE Trans. ASSP*, vol. 24, no. 4, pp. 320–327, 1976.
[19] T. Ueda *et al.*, in *Proc. Autumn Meeting of Acoustical Society of Japan*, 2022, pp. 165–168.
[20] D. R. Hunter *et al.*, *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
[21] T. Taniguchi *et al.*, in *Proc. HSCMA*, 2014, pp. 107–111.
[22] L. Li *et al.*, in *Proc. ICASSP*, 2020, pp. 846–850.
[23] Y. Mitsui *et al.*, in *Proc. ICASSP*, 2018, pp. 746–750.
[24] R. Scheibler *et al.*, in *Proc. ICASSP*, 2018, pp. 351–355.
[25] A. Kurematsu *et al.*, *Speech communication*, vol. 9, no. 4, pp. 357–363, 1990.
[26] E. Vincent *et al.*, *IEEE/ACM Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.