PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Detection of clean time-frequency bins based on phase derivative of multichannel signals

Atsushi HIRUMA[(1)], Kohei YATABE[(2)], Yasuhiro OIKAWA[(3)]

[(1)]Waseda University, Japan, y-sea.sc2dpt8h@ruri.waseda.jp
[(2)]Waseda University, Japan, k.yatabe@asagi.waseda.jp
[(3)]Waseda University, Japan, yoikawa@waseda.jp

Abstract

In this paper, a method for evaluating the cleanness of each time-frequency bin of multichannel spectrograms is proposed. When observing acoustical signals with noise and/or interference, the degree of noisiness is usually different for each bin in the time-frequency domain. Therefore, array signal processing techniques should be possible to be improved by choosing only "cleaner" bins, which contain less noise and/or interference, for extracting the spatial information. The proposed method aims to distinguish such clean bins from noisy ones. To do so, the similarity of phase derivative among channels is considered since phase is sensitive to noise and interference. Constant phase is removed by derivative so that convolutive mixtures can be handled without care on spatial condition. The proposed method is applied to direction-of-arrival estimation (MUSIC method) and blind source separation (independent vector analysis) for demonstrating the possibility of the proposed measure.

Keywords: Multichannel signal processing, microphone array, convolutive mixture, instantaneous frequency, group delay.

1 INTRODUCTION

Multichannel signal processing has been studied widely owing to its versatile applications [1–3]. The standard way of utilizing spatial information contained in the multichannel signals is to formulate the observation model in the time-frequency domain. In such formulation, each time-frequency bin represents the spatial information through phase difference and amplitude ratio between channels. Then, based on these information, the problem of handling multichannel signals is reduced to manipulation of a complex vector for each bin. In particular, phase difference between channels is important for signal processing because the plane wave model, which is considered in vast majority of multichannel signal processing techniques, does not admit amplitude difference but only phase contains its directional information.

When only a single source is observed without any noise or interference, then the spatial information can be easily extracted by just calculating the phase difference. However, such ideal situation cannot be realized in reality since every measurement is imperfect and contains noise to some extent. Moreover, many of the applications, including blind source separation (BSS) [4–6] and direction-of-arrival (DOA) estimation [7–9], begin with the situation where two or more sources are observed simultaneously (in such case, a source is interference for another sources). Therefore, estimating spatial information is far difficult in practical situation than the ideal situation. That is, one has to estimate it from phases contaminated by noise and/or interference.

To relieve this difficulty, it should be favorable to know the position of "cleaner" bins which contain less noise and/or interference. Since most of the interested signals (such as speech) admit sparse time-frequency representation, it can be expected that the degree of noisiness is different for each bin. That is, some bins are expected to be cleaner than the other bins. If one can know such cleaner bins in advance, it should be easier to estimate spatial information and obtain better processing result.

In this paper, we propose an indicator of cleanness of each time-frequency bin based on phase derivative of the multichannel signal. Since phase is sensitive to noise and interference, similarity of phase between channels should indicate the amount of noise and interference in each bin. However, phase difference also contains spatial information such as DOA, which does not allow to directly utilize the phase. To avoid such situation, derivative of phase is considered so that constant phase related to the spatial information is removed, and evaluation of

cleanness becomes possible. The potentiality of the proposed method is tested by DOA estimation and BSS, and its effectiveness is shown as the improvement of the performances.

2 MULTICHANNEL OBSERVATION MODEL

Let N source signals in time domain be written as N-dimensional vector $\mathbf{s}[\tau] = [s_1[\tau], s_2[\tau], ..., s_N[\tau]]^{\mathrm{T}}$, and M channel signals $\mathbf{x}[\tau] = [x_1[\tau], x_2[\tau], ..., x_M[\tau]]^{\mathrm{T}}$ be observed through convolution:

$$x_m[\tau] = \sum_{n=1}^{N} \sum_{k} h_{m,n}[k]\, s_n[\tau - k], \tag{1}$$

where $h_{m,n}$ is the impulse response relating nth source to the mth observation, and τ is the index of time. Through the Fourier transform, this relation can be represented in the frequency domain as

$$\hat{x}_m[\omega] = \sum_{n=1}^{N} \hat{h}_{m,n}[\omega]\, \hat{s}_n[\omega], \tag{2}$$

where ω is the index of frequency, and $\hat{h}$ denotes the Fourier transform of h. This frequency-domain representation is more convenient than that in the time domain as the convolution reduces to the element-wise multiplication. To gain such benefit while retaining the time-dependent information, Eq. (1) is often approximated by the short-time Fourier transform (STFT), which can be written as

$$\hat{\mathbf{x}}[t, \omega] \approx H[\omega]\, \hat{\mathbf{s}}[t, \omega], \tag{3}$$

where H is an $M \times N$ matrix whose (m, n)th element is $\hat{h}_{m,n}$, and t is another index of time in the time-frequency domain which depends on the setting of STFT.

The aim of this paper is to evaluate the degree of noisiness of observed signal $\hat{\mathbf{x}}[t, \omega]$ for each time-frequency index (t, ω) without assuming any structure on $H[\omega]$ or $\hat{\mathbf{s}}[t, \omega]$. That is, only inter-channel information at each bin is allowed for the evaluation, and any relation between time-frequency bins (such as harmonic structure and time continuity) cannot be utilized.

3 PROPOSED METHOD

For estimating cleanness of each time-frequency bin of observed multichannel signals, we propose to utilize derivatives of phase. Inter-channel difference of phase derivatives is measured by their standard deviation together with a nonlinear scaling function.

3.1 Phase of observed multichannel signal

For considering inter-channel information, the simplest situation is considered first. Let only nth source be active at (t, ω)th bin:

$$\hat{\mathbf{x}}[t, \omega] = \hat{\mathbf{h}}_n[\omega]\, \hat{s}_n[t, \omega], \tag{4}$$

where $\hat{\mathbf{h}}_n$ is nth column of H. That is, the observed signal $\hat{\mathbf{x}}$ is scalar multiple of $\hat{\mathbf{h}}_n$ which fully represents the spatial information between nth source and the m observed signals. Therefore, it should be beneficial to find such time-frequency bin for estimating or processing the spatial information.

In this paper, we consider fully blind setting for finding such "clean" time-frequency bin, i.e., no information on $H[\omega]$ or $\hat{\mathbf{s}}[t, \omega]$ is available. Then, it is impossible to distinguish $\hat{\mathbf{h}}_n[\omega]\, \hat{s}_n[t, \omega]$ from $\sum_n \hat{\mathbf{h}}_n[\omega]\, \hat{s}_n[t, \omega]$ because both $\hat{\mathbf{h}}_n$ and $\hat{s}_n$ are unknown. Therefore, the effect of either $\hat{\mathbf{h}}_n$ or $\hat{s}_n$ should be canceled so that the number of unknowns becomes manageable. Here, the effect of $\hat{\mathbf{h}}_n$ is canceled to reveal the information of source signals.

To do so, phase of the observed signal is considered. The polar representation of mth element of Eq. (4) can be written as

$$\hat{x}_m[t,\omega] = A_{\hat{h}_{m,n}}[\omega]\, \mathrm{e}^{j\varphi_{\hat{h}_{m,n}}[\omega]}\, A_{\hat{s}_n}[t,\omega]\, \mathrm{e}^{j\varphi_{\hat{s}_n}[t,\omega]}, \tag{5}$$

$$= \left(A_{\hat{h}_{m,n}}[\omega]\, A_{\hat{s}_n}[t,\omega]\right) \mathrm{e}^{j(\varphi_{\hat{h}_{m,n}}[\omega] + \varphi_{\hat{s}_n}[t,\omega])}, \tag{6}$$

where $j = \sqrt{-1}$, and $A_\cdot[t,\omega] \geq 0$. Therefore, its phase is given by

$$\varphi_{\hat{x}_m}[t,\omega] = \varphi_{\hat{h}_{m,n}}[\omega] + \varphi_{\hat{s}_n}[t,\omega]. \tag{7}$$

This representation of observed signal removes ambiguity on amplitude of $\hat{h}_{m,n}$. If the ambiguity on phase of $\hat{h}_{m,n}$ can also be eliminated, then information on the source signal can be extracted from the phase of observed signals. For instance, if $\varphi_{\hat{h}_{m,n}}[\omega]$ is constant for all m, then phases of all elements of $\hat{\mathbf{x}}[t,\omega]$ (indexed by m) are the same since $\varphi_{\hat{s}_n}[t,\omega]$ does not depend on m. That is, if $\varphi_{\hat{h}_{m,n}}[\omega]$ can be converted so that it becomes independent of m, then clean time-frequency bin should have the same phase for all channel m.

3.2 Partial derivatives of phase spectrograms

As the conversion of phase, its partial derivative is considered. Since (continuous) spectrogram depends on time and frequency, partial derivatives of phase are available for both time and frequency directions, which are called instantaneous frequency (IF) and group delay (GD), respectively [10–15]. Since derivative is linear, IF and GD of the clean time-frequency bin in Eq. (4) can be represented as

$$\mathrm{IF}_{\hat{x}_m}[t,\omega] = \partial_t \varphi_{\hat{x}_m}[t,\omega] = \partial_t \varphi_{\hat{h}_{m,n}}[\omega] + \partial_t \varphi_{\hat{s}_n}[t,\omega], \tag{8}$$

$$\mathrm{GD}_{\hat{x}_m}[t,\omega] = \partial_\omega \varphi_{\hat{x}_m}[t,\omega] = \partial_\omega \varphi_{\hat{h}_{m,n}}[\omega] + \partial_\omega \varphi_{\hat{s}_n}[t,\omega], \tag{9}$$

where each partial derivative can be computed either approximately or exactly [11] (note that their calculation is not expensive because computing all derivatives requires only three STFTs [11]).

These quantities are closely related to partial derivatives of the log-magnitude spectrogram [12]. If the Gaussian window is utilized for STFT, their relation can be written analytically as

$$\partial_t \varphi_{\hat{x}_m}[t,\omega] = C_1\, \partial_\omega \log|\hat{x}_m[t,\omega]| + C_2(\omega), \tag{10}$$

$$\partial_\omega \varphi_{\hat{x}_m}[t,\omega] = C_3\, \partial_t \log|\hat{x}_m[t,\omega]| + C_4(t), \tag{11}$$

where the coefficients C depend on the configuration of STFT, and these equations are valid for other window functions approximately. Since all the coefficients only depend on the STFT configuration (i.e., they do not depend on microphone index m), these equations indicate that inter-channel difference of IF and GD are small when the corresponding log-magnitude is smooth.

3.3 Two assumptions on phase derivatives of impulse response

Remind that $\varphi_{\hat{h}_{m,n}}[\omega]$ is phase of the impulse response in Eq. (1). Since log-magnitude spectrograms of usual room impulse responses can be expected to be smooth, we assume that inter-channel differences of $\partial_t \varphi_{\hat{h}_{m,n}}[\omega]$ and $\partial_\omega \varphi_{\hat{h}_{m,n}}[\omega]$ are small in practice. In addition, since the poles of typical impulse responses are independent of positions of sensors [16], we assume that $\partial_t \varphi_{\hat{h}_{m,n}}[\omega]$ and $\partial_\omega \varphi_{\hat{h}_{m,n}}[\omega]$ are similar for all channels. If at least one of these two assumptions is satisfied, then Eqs. (8) and (9) can be rewritten as

$$\mathrm{IF}_{\hat{x}_m}[t,\omega] \approx \partial_t \varphi_{\hat{s}_n}[t,\omega] + C, \tag{12}$$

$$\mathrm{GD}_{\hat{x}_m}[t,\omega] \approx \partial_\omega \varphi_{\hat{s}_n}[t,\omega] + C, \tag{13}$$

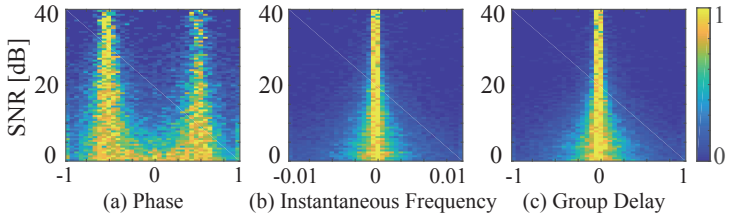


Figure 1. Histograms of inter-channel difference of (a) phase, (b) IF, and (c) GD. Each histogram was calculated from time-frequency bins which were categorized by bin-wise SNR of mixture of two female speech (dominating source was treated as S and the other was N for calculating SNR), and all histograms (normalized by maximum) were vertically concatenated to form the two-dimensional image.

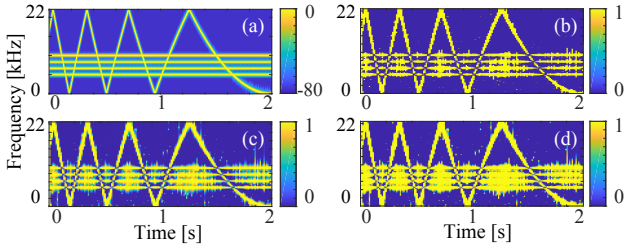


Figure 2. Illustrative example of the proposed measure of cleanness (calculated by IF) applied to (a) spectrogram of a synthetic signal. Each measure was calculated by (b) binary function in Eq. (16), (c) exponential function in Eq. (17), and (d) raised cosine in Eq. (18).

where C is some constant independent of m. That is, phase derivatives of clean time-frequency bins are the same for all channels.

This is illustrated in Fig. 1, where inter-channel difference of IF and GD was compared to that of phase (with frequency-wise phase alignment in [17]). Two-channel observation of two female speech signals was simulated so that they arrived from -30 and 30 degrees. For each time-frequency bin, inter-channel difference of (a) phase, (b) IF or (c) GD was calculated. At the same time, signal-to-noise ratio (SNR) of each bin was calculated by considering the larger source as signal and the other as noise. Normalized histograms of inter-channel difference of phase, IF or GD were calculated for each SNR, and stacked to form the two-dimensional images. As in the figure, inter-channel phase difference clearly indicates DOA of the source signals which was not perfectly concentrated but blurred for all SNR. In contrast, inter-channel difference of the phase derivatives was concentrated around 0, and it spread as SNR decrease. That is, difference of phase derivative of cleaner bins was nearly 0, while that of noisier bins had higher magnitude.

3.4 Proposed indicator of clean time-frequency bins

As in Eqs. (12) and (13), IF and GD of a clean time-frequency bin are expected to be independent of channel index m. That is, IF and GD are the same for all channel if that bin is clean. Therefore, similarity of IF and GD among channels is considered for detecting clean bins.

While a variety of similarity measures are available, standard deviation is considered in this paper. As the similarity is considered among channels, standard deviation is taken among channels as

$$\sigma_M(\hat{\mathbf{x}}[t,\omega]) = \left(\frac{1}{M} \sum_{m=1}^{M} \left(\partial_\cdot \varphi_{\hat{x}_m}[t,\omega] - \overline{\partial_\cdot \varphi_{\hat{\mathbf{x}}}}[t,\omega] \right)^2 \right)^{\frac{1}{2}}, \tag{14}$$

where $\partial_\cdot \varphi_{\hat{x}_m}[t,\omega]$ represents either $\mathrm{IF}_{\hat{x}_m}[t,\omega]$ or $\mathrm{GD}_{\hat{x}_m}[t,\omega]$ by denoting $\partial_\cdot \in \{\partial_t, \partial_\omega\}$, and $\overline{\partial_\cdot \varphi_{\hat{\mathbf{x}}}}[t,\omega]$ is its mean (taken for m). Based on the assumption in the previous subsection, a time-frequency bin having small

$\sigma_M(\hat{\mathbf{x}}[t,\omega])$ is expected to be cleaner than the higher ones. Since this quantity can be calculated for each (t,ω) independently, no structure within a channel is assumed.

For utilizing Eq. (14) in estimation and/or processing methods, it is convenient to apply a nonlinear scaling function which takes its value within 0 to 1. That is, the measure of cleanness is defined as

$$\mathcal{M}[t,\omega] = \mathcal{S}\big(\sigma_M(\hat{\mathbf{x}}[t,\omega])\big) \quad (\in [0,1]\,), \tag{15}$$

where $\mathcal{S}$ is the scaling function taking a real scalar between 0 and 1 such that $\mathcal{S}(0) = 1$, and $\mathcal{S}(\sigma)$ tends to 0 as σ increases. Although any function can be chosen for $\mathcal{S}$ to measure the similarity, three functions are defined in this paper for examples:

binary function taking either 0 or 1 decided by a threshold ε,

$$\mathcal{S}_{\text{binary}}(\,\sigma_M) = \begin{cases} 1 & (\,\sigma_M \le \varepsilon\,), \\ 0 & (\,\text{otherwise}\,), \end{cases} \tag{16}$$

exponential function with scaling parameter α,

$$\mathcal{S}_{\text{exp}}(\,\sigma_M) = \exp(-\alpha\,\sigma_M), \tag{17}$$

raised cosine of one period with scaling parameter α,

$$\mathcal{S}_{\cos}(\,\sigma_M) = \begin{cases} (\,1+\cos(\,\alpha\,\sigma_M))\,/\,2 & (\,|\,\alpha\,\sigma_M| \le \pi\,), \\ 0 & (\,\text{otherwise}\,). \end{cases} \tag{18}$$

By inserting one of these functions (or any other similar function) into Eq. (15), $\mathcal{M}[t,\omega]$ becomes a measure of cleanness of (t,ω)th bin, where $\mathcal{M}[t,\omega]$ is close to 1 if that bin is clean and close to 0 if noisy. The proposed measure with these scaling functions is illustrated in Fig. 2, where a two-channel signal was constructed from a sum of four sinusoidals and a chirp signal, and IF was used for the calculation of σ_M (using GD obtained quite similar results, and thus only IF is shown). As seen in the figure, the proposed measure takes 1 for time-frequency bins consisting of single component and takes nearly 0 for bins consisting of multiple components. Note that it takes nearly 0 for bins containing no component, which should be because their SNR is not high owing to the quantization error.

4 EXPERIMENTS

To demonstrate the usefulness of the proposed measure of cleanness, it was applied to three applications: random noise reduction by masking, DOA estimation by the multiple signal classification (MUSIC) method, and BSS by the independent vector analysis (IVA).

4.1 Simple random noise reduction by time-frequency masking

Since the proposed measure takes its value within 0 to 1, it can be regarded as a time-frequency mask if it correctly evaluates the degree of cleanness for each bin. To confirm the appropriateness of the proposed measure, a very simple denoising experiment was performed. A two-channel signal was simulated so that a single speech signal was arrived from 30 degree. The Gaussian random noise was added to each channel in the time domain so that its SNR becomes 0 dB, and the proposed measure was multiplied to the noisy signal in the time-frequency domain. The denoising results for each scaling function with several window lengths are summarized in Fig 3, where the horizontal axes represent parameters of the scaling functions. As SNR was improved for all situations, the results indicate that the proposed measure can correctly assign lower value for noisy time-frequency bins and higher value for the cleaner ones.

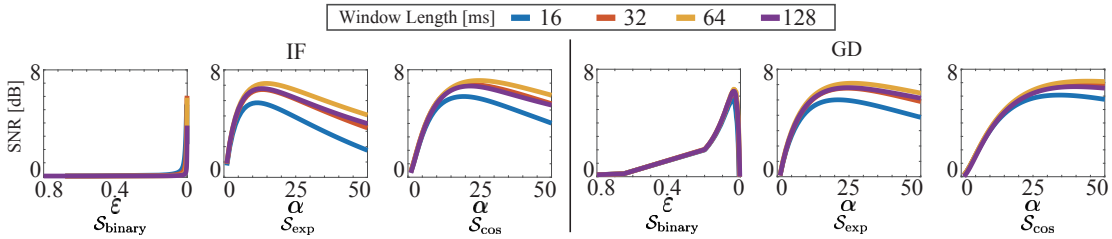


Figure 3. SNR improvement of random noise reduction by multiplying the proposed measure in the time-frequency domain. STFT was calculated by the half-overlapped Hann window with various window lengths, where the sampling frequency was 16 kHz.

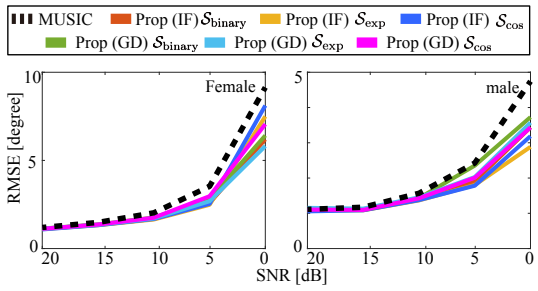


Figure 4. RMSE of estimated DOA by MUSIC method for each SNR.

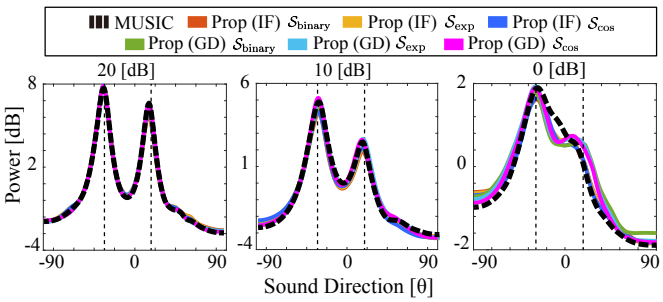


Figure 5. Examples of obtained MUSIC spectrums.

4.2 DOA estimation based on MUSIC spectrum

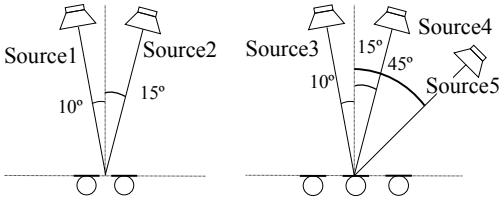
As an application of the proposed measure, DOA estimation using MUSIC method [18, 19] was considered. Observed signals were simulated by convoluting impulse responses to source speech signals which was obtained from SiSEC database [23]. The impulse responses convoluted to source signals were obtained from IKS database[1] whose T_{60} was 160 ms. True source angles were set to $\{-30, 15\}$ degrees, and the microphone spacing was 8 cm, where the distance between sound sources and microphones was 2 m. The Gaussian random noise was added to the simulated signals so that their SNR became $\{20, 15, 10, 5, 0\}$ dB, and the steering vectors were calculated by 0.1 degree increment between -90 to 90 degrees. The root-mean-squared error (RMSE) was calculated from 1000 random realizations. A single MUSIC spectrum was obtained for each realization by adding the spectrums for each frequency, and two largest peaks were chosen automatically from the peak positions obtained by `findpeaks` function in MATLAB. STFT was calculated by half-overlap Hann window whose length was 128 ms. The proposed measures were multiplied to the observed spectrograms before calculating the MUSIC spectrum, where the parameters were set to $\varepsilon = 0.1$ for $\mathcal{S}_{\text{binary}}$, $\alpha = 3$ for $\mathcal{S}_{\text{exp}}$, and $\alpha = 5$ for $\mathcal{S}_{\text{cos}}$.

RMSEs of estimated DOA are summarized in Fig. 4. While the result for MUSIC method with and without the proposed measure did not differ much in the cleaner situation, their difference is apparent in noisier situations. Multiplication of the proposed measure reduced RMSE in noisy situations, which can also be confirmed from Fig. 5 showing examples of MUSIC spectrum. In the case of 0 dB in Fig. 5, the smaller peak was resolved only when the proposed measure was multiplied, which is shown in color.

4.3 Determined BSS based on IVA

As another application, BSS using IVA [20–22] was considered. Live recordings of speech signals were obtained from SiSEC database [23] (`dev1`, `liverec`), where T_{60} was 130 ms, microphone spacing was 5 cm, and their geometry is shown in Fig. 6. After multiplying the proposed measure to the observed spectrograms, a set of

[1] https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/multi-channel-impulse-response-database/



(a) Two channel signals (b) Three channel signals

Figure 6. Observation settings of the BSS experiments.

Table 1. Separation results of two speech mixture [dB].

		SDR impr.		SIR impr.		SAR	
		female	male	female	male	female	male
IVA		1.9	7.7	4.3	11.2	7.3	10.5
Prop (IF) +IVA	$\mathcal{S}_{\text{binary}}$	2.2	7.6	4.6	11.2	7.3	10.4
	$\mathcal{S}_{\text{exp}}$	3.4	8.0	6.4	12.1	7.4	10.4
	$\mathcal{S}_{\text{cos}}$	3.6	8.0	6.5	12.2	7.6	10.4
Prop (GD) +IVA	$\mathcal{S}_{\text{binary}}$	**5.2**	7.7	8.4	11.4	**8.8**	10.4
	$\mathcal{S}_{\text{exp}}$	4.4	8.1	7.5	12.0	8.1	10.6
	$\mathcal{S}_{\text{cos}}$	5.1	**8.2**	**8.5**	**12.2**	8.3	**10.6**

Table 2. Separation results of three speech mixture [dB].

		SDR impr.		SIR impr.		SAR	
		female	male	female	male	female	male
IVA		1.1	4.0	3.6	6.9	4.4	5.7
Prop (IF) +IVA	$\mathcal{S}_{\text{binary}}$	2.4	5.1	5.6	8.2	4.6	6.0
	$\mathcal{S}_{\text{exp}}$	3.6	5.7	6.1	8.9	5.6	6.4
	$\mathcal{S}_{\text{cos}}$	2.9	**5.8**	5.9	9.1	5.0	**6.5**
Prop (GD) +IVA	$\mathcal{S}_{\text{binary}}$	**3.9**	3.8	**6.8**	6.9	**6.0**	5.6
	$\mathcal{S}_{\text{exp}}$	3.2	5.7	6.4	**9.4**	5.1	6.1
	$\mathcal{S}_{\text{cos}}$	3.2	5.2	6.6	8.8	4.7	5.8

demixing filters were obtained, and it was applied to the original observed signals. That is, multiplication of the proposed measure only affects the estimation of the demixing filter, and no noise reduction effect was retained in the separated results for fair comparison to the ordinary IVA. Separation results were evaluated by the Source-to-Distortion Ratio (SDR), Source-to-Interference Ratio (SIR), and Source-to-Artifacts Ratio (SAR) [24].

The results are summarized in Table 1 and Table 2. From the results, it can be seen that the best situations improved SDR about 3 dB for female speech of both two and three channel scenarios. Interestingly, the proposed method based on GD tended to be better than that based on IF. Note that applying the proposed measure improved the separation in most of the cases, and more importantly, it rarely degraded the separation performance. Therefore, the proposed method can be a good option for demixing matrix estimation in determined BSS. We emphasize again that these improvements were purely based on the improvement of demixing filters because the estimated filters were applied to the original observed signals.

5 CONCLUSION

In this paper, a method for evaluating the cleanness of each time-frequency bin of multichannel signals was proposed. By considering inter-channel difference of the partial derivatives of phase, the proposed method successfully distinguished clean time-frequency bins from noisy ones. Examples of applications on denoising, DOA estimation and BSS indicated the usefulness of the proposed measure in array signal processing. Although potentiality of the proposed measure was shown, its value is less understandable than the quantities utilized in sound enhancement literature, such as noise probability density functions and a priori SNR. Utilizing the proposed concept to obtain such understandable quantities should be the next step of research which remained as a future work.

REFERENCES

[1] M. S. Brandstein and H. F. Silvermana, “A practical methodology for speech source localization with microphone arrays,” Comput. Speech Lang., vol. 11, no. 2, pp. 91–126, 1997.

[2] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” Proc., vol. 60, no. 8, pp.

926–935, 1972.
[3] B. D. Van Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” IEEE Assp Mag., vol. 5, no. 2, pp. 4–24, 1988.
[4] 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, 2003.
[5] D. Kitamura, N. Ono, H. Ssawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factprization,” IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 24, no. 9, pp. 1626–1641, 2016.
[6] Y. Masuyama, K. Yatabe and Y. Oikawa, “Phase-aware harmonic/percussive source separation via convex optimization,” Proc. Int. Conf. Acoust. Speech Signal Process., pp. 985–989, 2019.
[7] L. C. Godara, “Application of antenna arrays to mobile communications. ii. beam-forming and direction-of-arrival considerations,” Proc., vol. 85, no. 8, pp. 1195–1245, 1997.
[8] S. Qin, Y. D. Zhang, and M. G. Amin, “Generalized coprime array configurations for direction-of-arrival estimation,” IEEE Trans. Signal Process., vol. 63, no. 6, pp. 1377–1390, 2015.
[9] T. Tachikawa, K. Yatabe, and Y. Oikawa, “Underdetermined source separation with simultaneous DOA estimation without initial value dependency,” Int. Workshop Acoust. Signal Enhanc., pp. 161–165, 2018.
[10] F. Auger and P. Flandrin, “Improving the readability of timefrequency and time-scale representations by the reassignment method,” Trans. Signal Process., vol. 43, no. 5, pp. 1068–1089, 1995.
[11] S. A. Fulop and K. Fitz, “Algorithm for computing the timecorected instantaneous frequency (reassigned) spectrogram,” J. Acoust. Soc. Am., vol. 119, no. 1, pp. 360–371, 2006.
[12] F. Auger and É. Chassande-Mottin and P. Flandrin, “On phasemagnitude relationships in the short-time fourier transform,” IEEE Signal Process. Lett., vol. 19, no. 5, pp. 267–270, 2012.
[13] K. Yatabe and Y. Oikawa, “Phase corrected total variation for audio signals,” Int. Conf. Acoust. Speech, Signal Process., pp. 656–660, 2018.
[14] K. Yatabe, Y. Masuyama, T. Kusano, Y. Oikawa, “Representation of complex spectrogram via phase conversion,” Acoust. Sci. Tech., vol. 40, no. 3, pp. 170–177, 2019.
[15] Y. Masuyama, K. Yatabe and Y. Oikawa, “Low-rankness of complex-valued spectrogram and its application to phase-aware audio processing,” Proc. Int. Conf. Acoust. Speech Signal Process., pp. 855–859, 2019.
[16] Y. Haneda, Y. Kaneda, and N. Kitawaki, “Common-acousticalpole and residue model and its application to spatial interpolation and extrapolation of a room transfer function,” IEEE Trans. Speech, Audio Process., vol. 7, no. 6, pp. 709–717, 1999.
[17] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” Signal Process., vol. 87, no. 8, pp. 1833–1847, 2007.
[18] R. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propagat., vol. 34, no. 3, pp. 276–280, 1986.
[19] S. A. Lee and K. Thomas, “A performance analysis of subspace-based methods in the presence of model errors, part i: The music algorithm,” IEEE Trans. Signal Process., vol. 40, no. 7, pp. 1758–1774, 1992.
[20] A. Hiroe, “Solution of permutation problem in frequency domain ICA, using multivariate probability density functions,” Proc. Indep. Comp. Anal., pp. 601–608, 2006.
[21] K. Yatabe and D. Kitamura, “Determined blind source separation via proximal splitting algorithm,” Int. Conf. Acoust. Speech, Signal Process., pp. 776–780, 2018.
[22] K. Yatabe and D. Kitamura, “Time-frequency-masking-based Determined BSS with Application to Sparse IVA,” Int. Conf. Acoust. Speech, Signal Process., pp. 715–719 2019.
[23] S. Araki, F. Nesta, E. Vincent, Z. Koldovskỳ, G. Nolte, A. Ziehe, and A. Benichoux, “The 2011 signal separation evaluation campaign (sisec2011):-audio source separation,” in Int. Conf. Latent Variable Anal. Signal Sep. Springer, pp. 414–422, 2012.
[24] V. Emmanuel, R. Gribonval and C. Févotte, “Performance measurement in blind audio source separation,” IEEE Trans. Audio Speech Lang. Process., vol. 14, no. 4, pp. 1462–1469, 2006.