Estimating Phase Linearity in The Frequency-Domain ICA Demixing Matrix

Keisuke Toyama[1,2] and Mark D. Plumbley[2]

[1] Sony Corporation, 1-7-1 Konan, Minato-ku, Tokyo, 108-0075, Japan
[2] Queen Mary University of London, Centre for Digital Music, Mile End Road, London, E1 4NS, United Kingdom
keisuke.toyama@jp.sony.com

Abstract. We consider a method for solving the permutation problem in blind source separation (BSS) by the frequency-domain independent component analysis (FD-ICA) by using phase linearity of FD-ICA demixing matrix. However, there is still remaining issue how we can estimate the phase linearity. In this paper, we propose two methods to estimate the linearity of the phase response of the FD-ICA demixing matrix. Our experimental result shows that our new methods can provide better estimation of the phase linearity than our previous method.

1 INTRODUCTION

Frequency-domain Independent Component Analysis (FD-ICA) [1] is a blind source separation method for convolutive mixtures, where separation using ICA is performed in the frequency domain, separately for each frequency component. However, successful use of FD-ICA involves solving a permutation problem, since the extracted sources in different frequency bins may be permuted relative to those in other frequency bins. The permutation problem can be tackled by using the direction of arrival (DOA) of each source [2, 3], but this can suffer from a spatial aliasing problem above a certain frequency limit. It is possible to deal with this problem by using phase linearity. Sawada et al. have proposed a method to solve this problem by estimating mixing model parameters and fitting to an idealised direct path mixing model, introducing a linear phase assumption on the FD-ICA mixing matrix [4]. We have proposed a method which uses linearity of the phase response of the demixing matrix directly [7, 8]. Nesta et al. also have proposed a similar method which uses the phase linearity based on time difference of arrival (TDOA) [6]. However, there is remaining issue how we should estimate the phase linearity. In this paper we propose new methods to estimate the phase linearity based on the DOA method or the Best Fit method.

2 BSS FOR CONVOLUTIVE MIXTURES

In the time-frequency domain, the observed signals at microphones $X_l(f,t)$ are expressed as

$$X_l(f,t) = \sum_{k=1}^{K} H_{lk}(f) S_k(f,t), \qquad l = 1, ..., L \tag{1}$$

where f represents frequency, t is the frame index, $H_{lk}(f)$ is the frequency response from source k to microphone l, and $S_k(f,t)$ is a time-frequency-domain representation of a source signal. Equation (1) can also be expressed as $\mathbf{X}(f,t) = \mathbf{H}(f)\mathbf{S}(f,t)$ where $\mathbf{X}(f,t) = [X_1(f,t), ..., X_L(f,t)]^T$ is the observed signal vector, $\mathbf{S}(f,t) = [S_1(f,t), ..., S_K(f,t)]^T$ is the source signal vector, and

$$\mathbf{H}(f) = \begin{bmatrix} H_{11}(f) & \cdots & H_{1K}(f) \\ \vdots & \ddots & \vdots \\ H_{L1}(f) & \cdots & H_{LK}(f) \end{bmatrix} \tag{2}$$

is the complex-valued mixing matrix.

In frequency-domain ICA, we perform signal separation from $\mathbf{X}(f,t)$ separately at each frequency f using the complex-valued demixing matrix

$$\mathbf{W}(f) = \begin{bmatrix} W_{11}(f) & \cdots & W_{1L}(f) \\ \vdots & \ddots & \vdots \\ W_{K1}(f) & \cdots & W_{KL}(f) \end{bmatrix} \tag{3}$$

which is adopted so that the reconstructed output signals $\mathbf{Y}(f,t) = [Y_1(f,t), ..., Y_K(f,t)]^T = \mathbf{W}(f)\mathbf{X}(f,t)$ become mutually independent. This can be done using any suitable ICA algorithm, such as the natural gradient approach [1]. Hereafter, we suppose we have two sources ($K = 2$) and two microphones ($L = 2$) for simplicity.

3 PERMUTATION PROBLEM

Since the ICA method has been applied separately at each frequency f, FD-ICA has an ambiguity in the order of the rows of $\mathbf{W}(f)$, such that permuted matrix is also the solution for FD-ICA. This problem is called as the *permutation problem* [1]–[3]. Methods designed to solve the permutation problem include the use of the amplitude correlation between adjacent frequencies [1, 3], and the use of the direction of arrival (DOA) [2, 3].

In the DOA method, we suppose a signal with frequency f comes from a source in the direction of θ. When the signal $\exp(j2\pi ft)$ is observed at the middle point of the microphones, the observed signals at the microphones are $X_l(f,t) = \exp\left(j2\pi f\left[t - d_l\sin(\theta_k(f))/c\right]\right)$, where d_l is the position of the microphone ($d_1 = -d_2 = D/2$) and c is the speed of sound. The frequency response of the demixing process between the observed signals and the separated signals is expressed by their ratio, $Y_k(f,t)/\exp(j2\pi ft)$. Thus, we can obtain the gain of the frequency response with respect to the direction as

$$\begin{aligned} G_k(\theta_k(f)) &= |Y_k(f,t)/\exp(j2\pi ft)| \\ &= |W_{k1}(f)\exp(-j2\pi f(d_1\sin(\theta_k(f)))/c) \\ &\quad + W_{k2}(f)\exp(-j2\pi f(d_2\sin(\theta_k(f)))/c)|. \end{aligned} \tag{4}$$

If $f < c/2D$, the gain $G_k(\theta_k(f))$ has at most one peak and one null point in a half period of $\theta_k(f)$ where $|\theta_k(f)| \le \pi/2$. The direction where the gain has the unique minimum value (null point) could be regarded as the direction of the unwanted source signal. Therefore, we can solve the permutation problem by comparing the direction of the two sources, $\theta_1(f)$ and $\theta_2(f)$. For more details of this process see [2, 3].

However, if $f > c/2D$, the gain $G_k(\theta_k(f))$ has two or more local minimum points so that we cannot uniquely determine the magnitude relationship between $\theta_1(f)$ and $\theta_2(f)$: this problem is called the *spatial aliasing problem.* For example, if the distance between two of microphones is 4 cm and the speed of sound is 343 m/sec, the spatial aliasing problem occurs for $f > 4288$ Hz.

However, by considering phase instead of direction or delay, we can obtain a new insight into this problem allowing us to reduce the spatial aliasing problem. A recent approach [4] to solve the permutation problem with the spatial aliasing problem is to estimate phase and amplitude parameters of an estimated mixing matrix $\hat{\mathbf{A}}(f) = \mathbf{W}^{-1}(f)$ assuming an anechoic direct path model. We have also proposed a method which uses the phase parameters, but for the *demixing* matrix $\mathbf{W}(f)$ directly [7, 8].

The direction of arrival $\theta_k(f)$ can be calculated as [3]:

$$\theta_k(f) = \arcsin\left(\frac{(\phi_k(f) - (2n_k(f)+1)\pi)c}{2\pi f D}\right) \tag{5}$$

where

$$\phi_k(f) = \angle W_{k1}(f) - \angle W_{k2}(f) \tag{6}$$

and $n_k(f)$ is an arbitrary integer to be determined such that $|(\phi_k(f) - (2n_k(f) + 1)\pi)c/2\pi f D| \leq 1$ is satisfied. However, if we plot the phase difference $\phi_k(f)$ itself, this often has an approximate linearity corresponding to constant delay. Thus, the difference could be represented by the following equation:

$$\hat{\phi}_k(f) = a_k f + b_k \tag{7}$$

where $b_k = \pm\pi$ and the equation holds modulo 2π. We know $b_k = \pm\pi$ since the DC component ($f = 0$) does not have phase information so that the two signals at two microphones should have opposite sign to suppress the signal. Our proposed method utilises this linear phase property. To solve the permutation problem, we estimate the a_k in (7) and then we calculate the distance between $\phi_k(f)$ and $\hat{\phi}_k(f)$.

4 PROPOSED ESTIMATION METHODS

In our previous papers [7, 8], we have estimated the parameters a_k and b_k by using the method of least squares on low frequency data where the spatial aliasing problem has not occurred ($f < c/2D$). However, the microphone spacing becomes wider, we can use fewer data values and the accuracy of the estimation of those parameters is likely to become worse. To prevent this problem, we propose new two methods to estimate those parameters.

4.1 Method based on DOA

To estimate the parameter a_k by using the method of least squares, we need reliable data where the permutation problem has not occurred. In a real environment, the DOA method is the most prevalent tool for solving the permutation problem in low frequencies where the spatial aliasing problem has not occurred [2, 3]. We therefore first apply the DOA method at low frequencies, then we use the phase linearity property to extend to higher frequencies by the following steps.

[Step 1] Solve the permutation problem by using the DOA method in low frequencies where the spatial aliasing problem has not occurred. Set initial loop counter $l := 1$.

[Step 2] Estimate a_k in (7) by using the method of least squares, as

$$a_k = \frac{\sum_{f \in \mathcal{F}} f \phi_k(f) - b_k \sum_{f \in \mathcal{F}} f}{\sum_{f \in \mathcal{F}} f^2} \tag{8}$$

$$b_k = \begin{cases} \pi & \text{if } \sum_{f \in \mathcal{F}_{low}} \phi_k(f) > 0 \\ -\pi & \text{otherwise} \end{cases} \tag{9}$$

where $\mathcal{F} = \{f : f_{low} \leq f \leq f_{high}^{(l)}\}$, $\mathcal{F}_{low} = \{f : f_{low} \leq f \leq c/2D\}$, and the frequencies f_{low} and $f_{high}^{(l)}$ are the low and high limits of the frequency range used to estimate a_k. For example, f_{low} is chosen to avoid the effect of low frequencies such as bins 5–20 [3]. $f_{high}^{(l)}$ is calculated at the Step 8. For the first loop, $f_{high}^{(1)} = c/2D$.

[Step 3] Estimate the lines $\hat{\phi}_k(f)$ (equation (7)).

[Step 4] Wrap the values of $\phi_k(f)$ and $\hat{\phi}_k(f)$ into $-\pi$ to π.

[Step 5] Calculate the distance $D_{prop}(f)$ between $\phi_k(f)$ and $\hat{\phi}_k(f)$ using

$$D_{prop}(f) = [E_{11}(f) + E_{22}(f)] - [E_{12}(f) + E_{21}(f)] \tag{10}$$

where

$$E_{ij}(f) = \begin{cases} |\phi_i - \hat{\phi}_j| & \text{if } |\phi_i - \hat{\phi}_j| < \pi \\ 2\pi - |\phi_i - \hat{\phi}_j| & \text{otherwise.} \end{cases} \tag{11}$$

[Step 6] Solve the permutation problem by using $D_{prop}(f)$. If $D_{prop}(f) < 0$, consider that a permutation has occurred at the frequency f, whereas if $D_{prop}(f) > 0$, a permutation has not occurred at the frequency f.

[Step 7] Calculate the set of "phase wrapping" frequencies $\mathcal{F}_{wrap}$ as

$$\mathcal{F}_{wrap} = \{0 < f \leq f_{\max} : \hat{\phi}_k(f) = \pm(2n+1)\pi, \;\; k = 1, 2\} \tag{12}$$

where $f_{\max}$ is the Nyquist frequency. If no wrapping frequencies exist, $\mathcal{F}_{wrap}$ is a null set.

[Step 8] Make the set of the high limit frequencies to estimate the a_k as

$$\{f_{high}^{(l)}\} = \{c/2D\} \cup \mathcal{F}_{wrap} \cup \{f_{\max}\}. \tag{13}$$

The $(l{+}1)$-th smaller number is used as $f_{high}^{(l)}$ at the Step 2 in the next loop.

[Step 9] Unwrap the values of ϕ_k as

$$\phi_k(f) = \phi_k(f) + \mathrm{sign}(a_k) 2\pi m_k \tag{14}$$

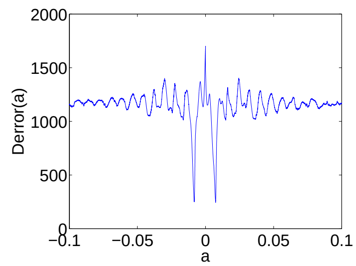


Fig. 1. a and $D_{error}(a)$.

where m_k is chosen to keep the line continuous.
[Step 10] Increase the loop counter $l := l + 1$ and update the set of frequencies $\mathcal{F}$ as $\mathcal{F} = \{f : f_{low} \leq f \leq f_{high}^{(l)}\}$, and then repeat from the Step 2. In the final loop when $f_{high}^{(l)} = f_{\max}$, do the Step 2–6 and then stop.

4.2 A "Best fit" method

The method which is described in Section 4.1 needs the observed data for which the permutation problem has already been solved using e.g. the DOA method. Thus, the performance of the estimation of the parameters a_k depends on the performance of the initial permutation solver. Here, we propose a new method which does not require these conditions.

To estimate the a_k, we calculate the distance between ϕ_k and assumed linear curves calculated by using all possibility of the value of a_k. We assume the linear curve $\tilde{\phi}(a, f)$ as

$$\tilde{\phi}(a, f) = af + b \tag{15}$$

$$b = \begin{cases} \pi & \text{if } \ a < 0 \\ -\pi & \text{otherwise} \end{cases} \tag{16}$$

where a is assumed value and the range of the value depends on the location of sources and microphone. Here, we use $\{a : -0.1 < a < 0.1\}$. Next, we calculate the distance $D_{error}(a)$ between $\tilde{\phi}(a, f)$ and $\phi_k(f)$ for all values of a.

$$D_{error}(a) = \sum_f \min(|\tilde{\phi}(a, f) - \phi_1(f)|, |\tilde{\phi}(a, f) - \phi_2(f)|) \tag{17}$$

The distance $D_{error}(a)$ has two local minimum values (see the Figure 1). These local minimum values should be the estimated value of a_k. To solve the permutation problem, we calculate the distance between $\phi_k(f)$ and $\hat{\phi}_k(f)$ as same as the Step 5 and 6 in Section 4.1.

5 EXPERIMENTS

To confirm our methods, we performed experiments to separate two speech signals (5 sec of speech at 44.1 kHz) mixed using impulse responses of an anechoic room ($T_{60} = 0$ msec), an echo room ($T_{60} = 300$ msec), a Japanese Tatami floored

Table 1. Error rate obtained with the inter-frequency correlation method, the DOA method, and the proposed methods.

Error rate [%]	Inter-freq. correlation	DOA	Proposed A	Proposed B	Proposed C
Anechoic Room	30.15	0.88	3.41	8.78	8.39
Echoic Room A	6.83	60.59	15.12	4.59	6.44
Japanese Tatami Room	28.78	12.68	41.95	8.68	7.02
Conference Room	34.93	54.44	17.56	9.17	6.63

room (T_{60} = 600 msec), and a conference room (T_{60} = 780 msec). These impulse responses are supplied by RWCP database. The distance between the two microphones is 2.83 cm, so spatial aliasing would begin at 6060 Hz. For the FD-ICA part, we adopt 2048 as the length of FFT window, and run for 300 iterations. In these experiments, we compared the performance of our methods (method A described in [7], method B described in Section 4.1, and method C described in Section 4.2) to the inter-frequency correlation method [1, 3] and the DOA method [2, 3]. For the proposed methods, we used 5 as the lowest frequency bin number f_{low}.

Here, we define "correct" permutation data to evaluate the performance of the inter-frequency correlation method and our proposed methods. The "correct" data are obtained by the correlation between the input signal $U_{lk}(f,t) = H_{lk}(f)S_k(f,t)$ observed at microphone and the separated signal $Z_{lk}(f,t) = W_{lk}^{-1}(f)Y_k(f,t)$ which is projected to the microphone by the inverse matrix of the demixing matrix at each frequency [8].

The results are shown in Figures 2 and 3, and Tables 1 and 2. The methods B and C can solve the permutation better than the inter-frequency correlation method and the DOA method in all the environments. It can be seen from the Figure 3, the estimation of phase linearity is better in methods B and C than for method A which uses only low frequency data to estimate the parameter a_k, especially in the Japanese Tatami Room. It seems that the estimation of the linear curve of method C is slightly better than that of method B. Because the method B relies on the performance of previous permutation solver in each loop for estimating the linear curve, so that the estimation error would be accumulated. On the other hand, method C estimates the linear curve directly from the observed data. Thus, the method C does not suffer from such error essentially. However, if we do not know the direction of sources, we have to calculate all direction for estimating the linear curve.

6 CONCLUSION

We have proposed two methods which estimate the linearity of the phase response of the demixing matrix to solve the permutation problem. While the permutation errors for methods B and C are significantly reduced from method A, the SIR (Table 2) is only slightly reduced, probably due to most of the energy in these signals being in the lower frequency regions. The proposed methods can estimate the phase linearity better than inter-frequency correlation method,

Table 2. Comparison of average SIR [5] obtained with the inter-frequency correlation method, the DOA method, and the proposed methods. All values are expressed in decibels (dB).

SIR [dB]	Inter-freq. correlation	DOA	Proposed A	Proposed B	Proposed C
Anechoic Room	23.860	22.478	32.725	32.725	32.542
Echoic Room A	17.538	16.037	17.543	17.543	17.549
Japanese Tatami Room	2.346	3.411	3.451	3.451	3.455
Conference Room	0.191	3.234	3.378	3.378	3.406

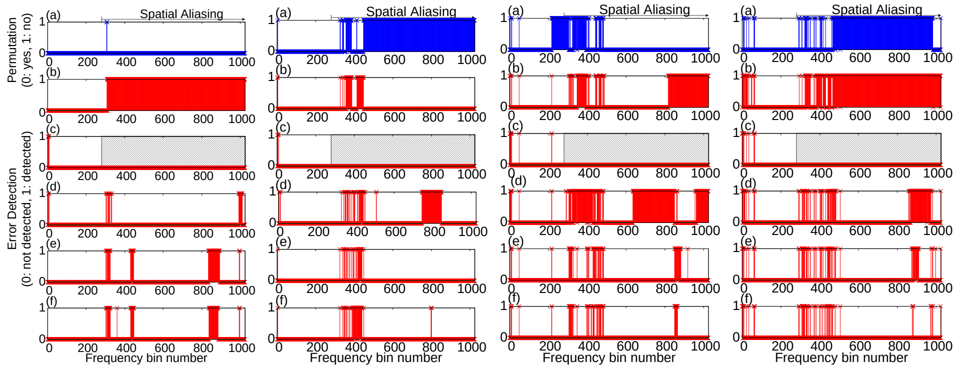
the DOA method, and our previous method in [7, 8] especially in long reverberant environments. In future work, we plan to compare the performance of our methods with that of Sawada's method [4] and Nesta's method [6].

7 ACKNOWLEDGES

We used the RWCP Sound Scene Database in Real Acoustical Environments. We would like to thank associate professor Takanobu Nishiura of Ritsumeikan University for giving us useful comments about the database. Mark D. Plumbley is supported by an EPSRC Leadership Fellowship (EP/G007144/1).

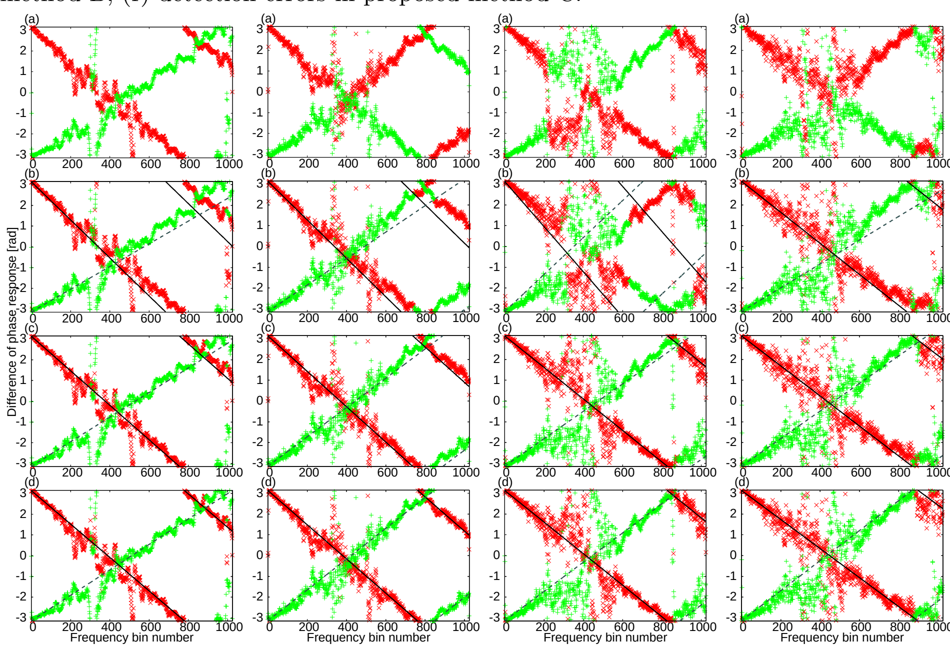
References

1. P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp.21–34, 1998.
2. S. Kurita, *et al*, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP 2000*, vol. 5, pp. 3140–3143, 2000.
3. H. Sawada, *et al*, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transaction. Speech and Audio Proc.*, vol. 12, pp. 530–538, 2004.
4. H. Sawada, *et al*, "Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation," *IEEE Trans. Audio, Speech, and Language Proc.*, vol. 15, pp. 1592–1604, 2007.
5. E. Vincent, *et al*, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 14, pp.1462–1469, 2006.
6. F. Nesta, *et al*, "A Novel Robust Solution to the Permutation Problem Based on a Joint Multiple TDOA Estimation," *Proc. IWAENC 2008*, paper ID 9023, 2008.
7. K. Toyama, *et al*, "Using phase linearity to tackle the permutation problem in audio source separation," *Proc. IWAENC 2008*, paper ID 9040, 2008.
8. K. Toyama, *et al*, "Solving the permutation problem using phase linearity and frequency correlation," *Proc. ICArn International Workshop 2008*, pp.48–51, 2008.



(i) Anechoic Room (ii) Echoic Room (iii) Japanese Tatami Room (iv) Conference Room

Fig. 2. Detection of permutation in (i) the anechoic room, (ii) the echoic room, (iii) the Japanese Tatami room, (iv) the conference room; (a) correct detection, (b) detection errors in the inter-frequency correlation method, (c) detection errors in the DOA method, (d) detection errors in proposed method A, (e) detection errors in proposed method B, (f) detection errors in proposed method C.



(i) Anechoic Room (ii) Echoic Room (iii) Japanese Tatami Room (iv) Conference Room

Fig. 3. The demixing matrix phase response in the same rooms as Fig. 2, showing observed points ϕ_1 ('×'), ϕ_2 ('+'), and estimated lines $\hat{\phi}_1$ (solid line), $\hat{\phi}_2$ (dashed line) from equations (6) and (7). In each environment, (a) shows ϕ_k before solving the permutation problem, (b)–(d) show ϕ_k and $\hat{\phi}_k$ after solving the permutation using method A (b), method B (c), and method C (d).