

USING PHASE LINEARITY IN FREQUENCY-DOMAIN ICA TO TACKLE THE PERMUTATION PROBLEM

Keisuke Toyama^{†‡} and Mark D. Plumbley[‡]

[†] Sony Corporation, 1-7-1 Konan, Minato-ku, Tokyo, 108-0075, Japan

[‡] Queen Mary University of London, Centre for Digital Music, Mile End Road, London, E1 4NS, United Kingdom

keisuke.toyama@jp.sony.com

ABSTRACT

This paper describes a method for solving the permutation problem in the frequency-domain independent component analysis (FD-ICA) approach to blind source separation (BSS). FD-ICA is a well-known method for BSS of convolutive mixtures. However, FD-ICA has a source permutation problem, where estimated source components can become swapped at different frequencies. Many researchers have suggested methods to solve the source permutation problem including using correlation between adjacent frequencies and direction of arrival (DOA). We propose a modification to the DOA method, based on phase linearity of the FD-ICA de-mixing matrix, that can extend the range of frequencies over which the permutation problem can be resolved. Experiments indicate that our method can provide a better performance than the inter-frequency correlation method and the DOA method in real environments.

Index Terms— Independent component analysis (ICA), permutation problem, direction of arrival (DOA), linearity, phase response.

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 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. Recently, Sawada et al. [4] have proposed to ameliorate the latter 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. In this paper we investigate an alternative method exploiting this idea, based on a linear phase property of the de-mixing matrix directly.

2. BSS FOR CONVOLUTIVE MIXTURES

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

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

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 $\mathbf{X}(f, t) = [X_1(f, t), \dots, X_L(f, t)]^T$ is the observed signal vector, $\mathbf{S}(f, t) = [S_1(f, t), \dots, S_K(f, t)]^T$ is the source signal vector, and

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

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 de-mixing matrix

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

so that the reconstructed output signals $\mathbf{Y}(f, t) = [Y_1(f, t), \dots, Y_K(f, t)]^T = \mathbf{W}(f)\mathbf{X}(f, t)$ 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 $X_l(f, t) = \exp(j2\pi f[t - d_l \sin(\theta_k(f))/c])$, 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 de-mixing 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

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

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 $|\theta_k(f)| \leq \pi/2$. 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. In contrast, we propose a new method which uses the phase parameters for the *de-mixing* matrix $\mathbf{W}(f)$ directly.

4. PROPOSED METHOD

According to [3], the direction of arrival $\theta_k(f)$ can be calculated as:

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

where

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

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 (Fig. 1), we see that the difference has a linearity corresponding to approximately constant delay. Thus, the difference could be represented by the following equation [6]:

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

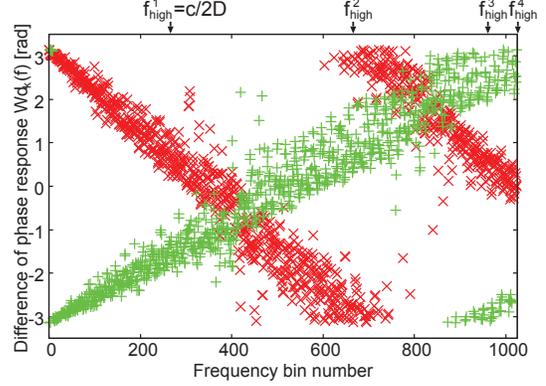


Fig. 1. Plot of the observed de-mixing matrix phase difference ϕ_k in an echoic environment showing ϕ_1 ('x') and ϕ_2 ('+'). $f_{high}^{(l)}$ show the high limits of the frequency range used to estimate parameters in equation (13).

where $b_k = \pm\pi$ and the equation holds modulo 2π . We know $b_k = \pm\pi$ since the DC component 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 linear curves and then we calculate the distance between $\phi_k(f)$ and $\hat{\phi}_k(f)$. Sawada et al. [4] have made a similar observation on the phase difference of the estimated inverse matrix of the de-mixing matrix $\mathbf{A}(f) = \mathbf{W}^{-1}(f)$. However, we do not need to calculate the inverse matrix in our method, but we just use the *de-mixing* matrix.

To estimate the parameters a_k and b_k , 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 and b_k in (7) by using the method of least squares, as

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

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

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 and b_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 between $\phi_k(f)$ and $\hat{\phi}_k(f)$,

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

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

[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

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

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 a_k and b_k as

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

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

$$\phi_k(f) = \hat{\phi}_k(f) + \text{sign}(a_k)2\pi m_k \quad (14)$$

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 Step 2. In the final loop when $f_{high}^{(l)} = f_{max}$, do the Step 2–6 and then stop.

While [4] uses the phase linearity of the estimated mixing matrix $\hat{\mathbf{A}}$ by calculating the inverse of the de-mixing matrix \mathbf{W} , our method uses the phase linearity of the de-mixing matrix \mathbf{W} , so we are steering the null towards the unwanted source. In this way, we hope to remove the direct path of the unwanted source.

5. EXPERIMENTS

To confirm our method, 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 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 method to the inter-frequency correlation method and the DOA method. For the proposed method, we used 5 as the lowest frequency bin number f_{low} to estimate a_k and b_k by an exploratory experiment.

Here, we define “correct” permutation data to evaluate the performance of the inter-frequency correlation

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

Error rate [%]	Inter-freq. correlation	DOA	Proposed
Anechoic Room	30.15	0.88	8.78
Echoic Room A	6.83	60.59	4.59
Japanese Tatami Room	28.78	12.68	8.68
Conference Room	34.93	54.44	9.17

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

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

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 de-mixing matrix at each frequency [7].

The results are shown in Figures 2 and 3, and Tables 1 and 2. The proposed method can solve the permutation better than the inter-frequency correlation method and the DOA method in all the environments. However, the proposed method can have errors at frequencies near where two estimated linear phase lines cross (e.g. frequency bins 431 and 862 in the anechoic room). Around those points, this method cannot distinguish two sources, because the two sources have the same phases at these points. As one possible solution to this problem, we proposed a combination method between our method and the inter-frequency correlation method [7]. However, the performance will depend on the result of the inter-frequency correlation method.

6. CONCLUSION

We have proposed a method which uses the linearity of the phase response of the de-mixing matrix to tackle the permutation problem in blind audio source separation. The proposed methods can yield better performance in frequency-domain ICA than that of the inter-frequency correlation method and the DOA method. However, the proposed method does have a difficulty around the points where the two linear curves of the difference of the phase response cross, and it cannot solve the permutation at those points.

While Sawada et al. [4] estimates the mixing process, the concept of our method fits the ICA concept of “notching out” unwanted sources by the de-mixing matrix.

In future work, we will be comparing our de-mixing matrix phase linear method with the mixing matrix parameters method of [4]. We are also considering combining with alter-

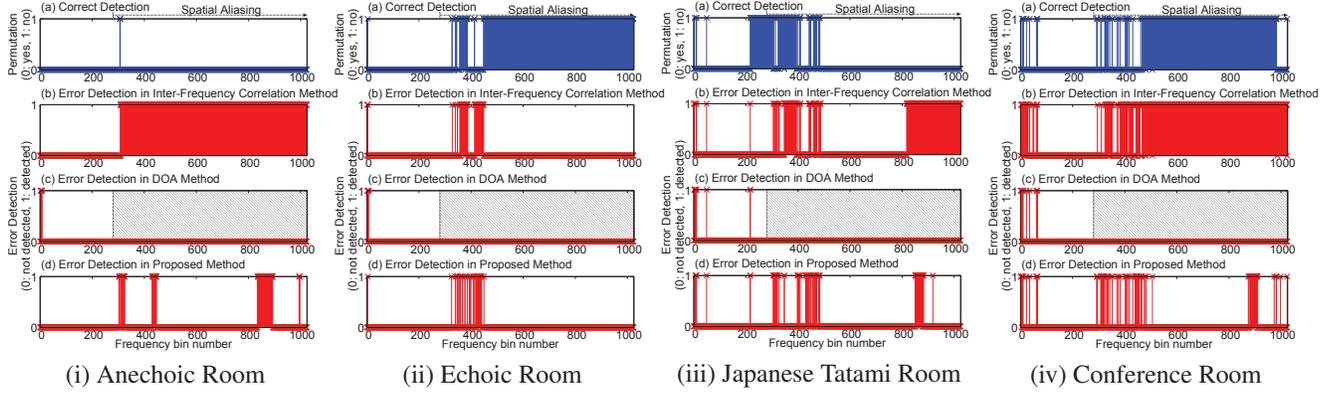


Fig. 2. Detection of permutation in (i) the anechoic room, (ii) the echoic room, (iii) the Japanese Tatami room, (iv) the conference room; (a) correct detection, (b) detection errors in the inter-frequency correlation method, (c) detection errors in the DOA method, (d) detection errors in proposed method.

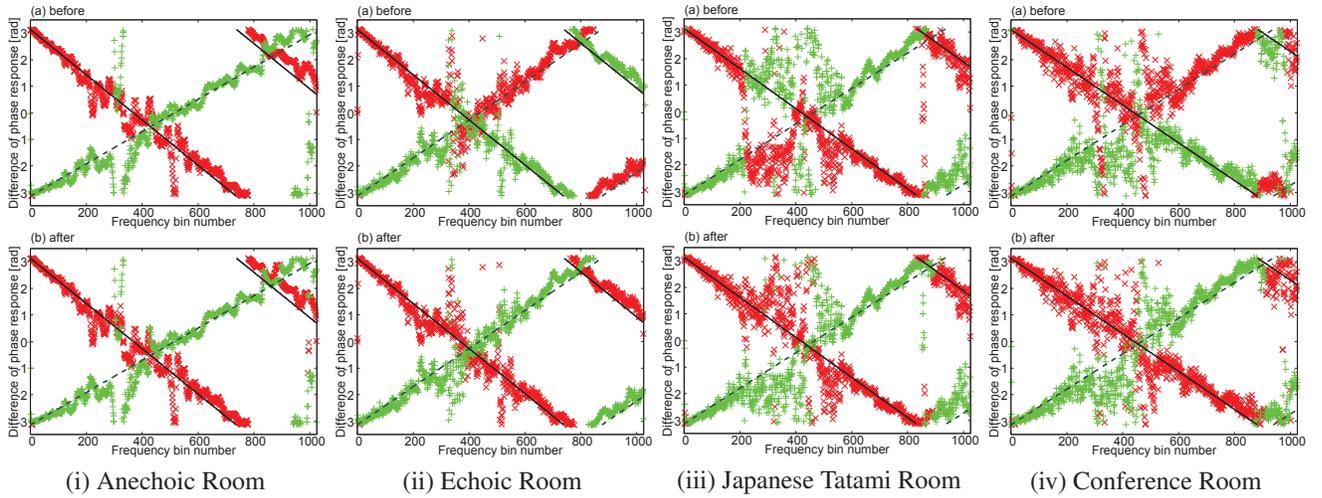


Fig. 3. The de-mixing matrix phase response in the same rooms as Fig. 2, showing observed points ϕ_1 ('x'), ϕ_2 ('+'), and estimated lines $\hat{\phi}_1$ (solid line), $\hat{\phi}_2$ (dashed line) from equations (6) and (7).

native methods based on amplitude information of sources to solve the permutation problem around those points.

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).

8. 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 Trans. 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] K. Toyama, *et al.*, "Using phase linearity to tackle the permutation problem in audio source separation," *Proc. IWAENC 2008*, paper ID 9040, 2008.
- [7] K. Toyama, *et al.*, "Solving the Permutation Problem using Phase Linearity and Frequency Correlation," *Proc. ICAn International Workshop 2008*, pp.48–51, 2008.