독립성분분석을 이용한 교환 모호성의 해소

김민제*, 장인선+, 백승권+, 강경옥+, 최승진*, 홍진우+

*포항공과대학교 컴퓨터공학과

+한국전자통신연구원 디지털방송연구단 방송미디어연구그룹

Independent Component Analysis for Solving Permutation Ambiguity

M. Kim*, I. Jang+, S. Beack+, K. Kang+, S. Choi*, J. Hong+

*Department of Computer Science, Pohang University of Science and Technology (Postech)

*{minjekim, seungjin}@postech.ac.kr

+Digital broadcasting Research Division, Broadcasting Media Research Group,

Electronics and Telecommunications Research Institute (ETRI)

+{jinsn, skbeack, kokang, jwhong}@etri.re.kr

Abstract

Permutation ambiguity has been a bottleneck of frequency domain method for convolutive blind separation. We focused on the fact that there are frequency dependent factors in the (nearfield) mixing filter model and that can be reduced by sophisticated normalization. After that, the only remaining dependency factors of mixing filter model are positions of sources and sensors which can give some geometric information of basis vectors of mixing filter model. We can finally use the information to cluster the permuted basis vectors. Proposed method can find out the geometric information using independent component analysis without the normalization steps. Hence it can be an alternative when we have no prior information about the distance between sources and microphones, which are indeed input parameters of previous normalization steps. Experiments with comparing to other models, confirm the high performance of proposed method.

1. Introduction

If convolutive blind source separation (BSS) problem is solved by transforming the input time series signals into time-frequency domain and performing instantaneous technique separately at each frequency, we have to be confronted with permutation ambiguity problem.

There have been many researches to tackle this problem. Introducing constraints on filter models and smoothing the estimated frequency response can be a good solution to that, but it can not be used when the filter length is long [1] [2]. Or we can sort the permutation using the correlation between the envelopes of band-passed signals which is precise but not robust since misalignment at a frequency is propagated consecutive frequencies [3]. Another robust method is based on direction of arrival (DOA) [4], but in this case, the wavelength must be longer than the distance of sensors so that we can not use it when the sampling rate of observed signal is high. Finally, if the independence of estimated source signals is evaluated in the time domain, we do not have to worry about the permutation ambiguity but it can lose the benefit of frequency-domain BSS [5] [6].

We concentrated on the current work that exploits the information of basis vectors produced by independent component analysis (ICA) [7]. It needs the only prior knowledge, simply the maximum distance of a sensor and the other ones. Normalized basis vectors using the prior information can be clustered afterward and the clustering results are used to solve the permutation ambiguity. Our proposed method eliminates these constraints about the distance of sensors while it keeps the separation performance good. The main idea of our system is to find out the geometric information of mixing matrices using ICA again without any prior information about the mixing process.

The next section describes how we can find and use the geometric information of mixing process to order the permutation using sophisticated normalization steps. Section 3 shows our proposed method. Section 4 is devoted to experimental results and section 5 concludes our work.

2. Normalization of Basis Vectors [7]

A previous work showed that proper normalization of all the basis vectors can reduce their variation of frequency. That means when we approximate the multi-path convolutive mixing model by using a direct-path (nearfield) model

$$h_{jk}(f) \approx \frac{q(f)}{d_{jk}} \exp[i 2\pi f c^{-1}(d_{jk} - d_{Jk})], \qquad (1)$$

where f is the frequency and c is the velocity of signal and $d_{jk} > 0$ is the distance between source k and sensor j. Note that there are frequency dependent factors f and $q(f)$ so that they make the frequency response values $h_{jk}(f)$ vary with the change of frequency.

The specialized normalization consists of two steps. For all ICA basis vectors $\mathbf{a}_i(f)$, $i = 1, \ldots, M$, where M is the number of sources, we select a reference sensor J and then calculate

$$\bar{a}_{ji}(f) \leftarrow | a_{ji}(f) | \exp\left[i \frac{\arg[a_{ji}(f) / a_{Ji}(f)]}{4 f c^{-1} d_{\max}} \right] \qquad (2)$$

where $d_{\max}$ is the maximum distance between the reference sensor J and a sensor $\forall j \in \{1, \ldots, M\}$ and that is the only prior information of this normalization method. The second step is simple unit-norm normalization.

After the normalization, we can have the normalized frequency response value estimators

$$\bar{a}_{ji}(f) \approx \frac{1}{d_{jk} D} \exp\left[i \frac{\pi}{2} \frac{(d_{jk} - d_{Jk})}{d_{\max}} \right], \quad D = \sqrt{\sum_{i=1}^{M} \frac{1}{d_{ik}^2}}. \qquad (3)$$

The only remaining dependency is the positions of the sources and sensors so that the basis vectors $\mathbf{a}_i(f)$ forms M clusters. Our next step is to make the i th basis vectors $\mathbf{a}_i(f)$ in the same cluster found by general clustering methods such as k-means clustering. You can have more detailed description about this permutation ordering process in [7]. Fig. 1 shows the comparison between initial frequency response values and normalized ones. You can see that the normalization reduced the variations of frequency. Fig. 2 describes the real part of normalized basis vectors when the number of sensors and sources are all two.

3. ICA for Permutation Ordering

We found out the importance of the fact that the remaining dependency is the positions of the sources and sensors after the normalization. The main idea of previous work is that the frequency dependent factors should be eliminated to cluster the basis vectors and order the permutation. However, our first intuition is that we do not have to do the elimination because the remaining information, the distances of sources and sensors, is represented in the data as angles of crossed data points. You can see this fact in Fig. 2 that data points, the column vectors of

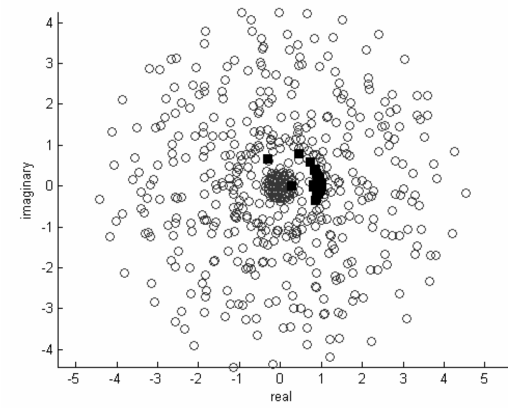


Fig. 1. *This figure shows the reduced dependency on frequency. Horizontal axis is the real part of frequency response values and vertical axis is their imaginary part. Empty circles represent the initial values and black squares correspond to the normalized ones.*

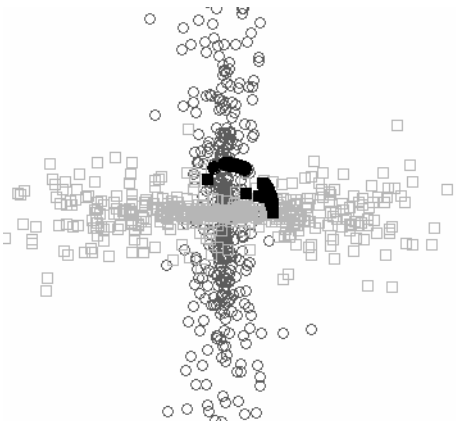


Fig. 2. *Real part of basis vectors and normalized ones. The initial empty circles and squares are reduced to black dots and black squares respectively compounding two clusters.*

the frequency response, make two rough lines crossing at the origin. If we normalize these vectors, we can have some clusters which contain the directions of the rough crossing lines, but if we can find out the two directions of vectors directly, we do not have to normalize and the results will be similar to the clustering results of normalized vectors.

For example, we can find out two independent components (ICs) from the frequency responses in previous examples. In this case the mixing filter transformed to frequency domain is $2 \times 2 \times L$ cubic matrix where L is the number of frequency bins and there are two sources and two sensors. So each L column vectors consists horizontal or vertical rough lines and the direction of each line can be found by ICA (See Fig. 3). We cluster these vectors based on their encoding values. For instance, if a vector is on the same direction of a unit-norm normalized IC, its corresponding encoding value is higher than

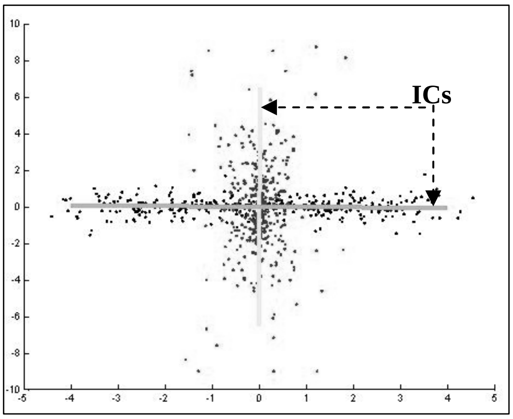


Fig. 3. *Column vectors make two rough crossing lines and the found ICs represents the direction of lines.*

the other encoding values which correspond to the other ICs.

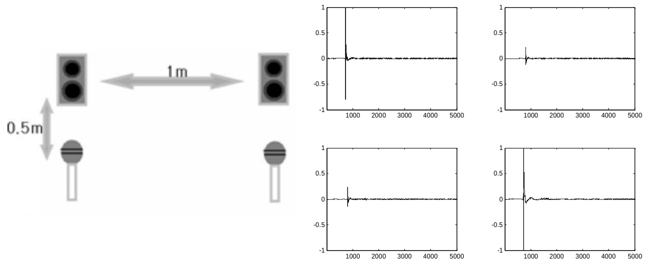
Our proposed method can overcome the constraints or weak points of previous normalization method. At first, the previous normalization method needs an input parameter $d_{\max}$ which is the maximum distance of a given reference sensor and the other ones while the proposed method does not need any prior knowledge. The second is about the selection of reference sensor. The performance of normalization is a little bit dependent on which sensor we select as a reference sensor. Without this selection step, our proposed method provides robust ordering results of permuted basis vectors.

4. Experimental Results

In this section we will first show the clustering performance of the normalization method and our proposed one using artificially permuted mixing filter and then the separation results will be given with their enhanced SIR values. The separation experiments use convolutive mixtures recorded in real environment.

4.1 Clustering Performance

We measured some impulse responses (see Fig. 4 (b)) in the environment given in Fig. 4 (a). Then we transformed these time-domain signals into frequency domain and got $2\times2\times512$ cubic mixing filter matrix. For a given frequency value f, there are two basis vectors $\mathbf{a}_1(f)$ and $\mathbf{a}_2(f)$. We have already seen that the basis $\mathbf{a}_1(f)$ and $\mathbf{a}_2(f)$ vectors consists vertical or horizontal intrinsic lines respectively. We artificially permuted these real mixing filers and ordered the permutation using normalization method and our proposed method (see Fig. 5) respectively. We used complex-valued FastICA [8] to find out the ICs from the 1024 basis vectors. Tab. 1 shows the clustering performance of the two methods in various environments. From this result, we can see that if each sensor has corresponding dominant source, permutation



(a) (b)

Fig. 4. *The mixing environment is shown in (a) and measured impulse responses are in (b). Each column in (b) represents impulse responses of each source measured by two sensors.*

ordering methods based on clustering will yield good performance. Another important point is that normalization method is a little bit sensitive to selection of reference sensors but our proposed method is not.

4.2 Separation Results

We checked our performance using speech data which are two mixtures of female voice and male voice recorded with 2 microphones in the real studio environment. At first we transform the mixture signals into time-frequency domain and then applied complex-valued ICA which uses the logistic infomax [9] with the natural gradient feature [10] and then we applied permutation ordering methods and finally solved scaling ambiguity by making dominant sensor responses 1 [7]. The sampling rate is 16kHz and the frame size is 1024 (64ms). The performance was evaluated in terms of the signal-to-interference ratio (SIR) improvement. We do not have channel response and underlying signals so that we used alternating signals in the same way proposed by [11]. The SIR

	ICA	Norm 1	Norm 2
Case 1	0.20	0.59	0.59
Case 2	3.71	4.88	5.08
Case 3	27.34	31.05	29.10

Tab. 1. *Case 1 is the mixing environment described in Fig. 4 (a). Case 2 is the same situation but the distance of sources and that of sensors are reduced to 0.5m. Case 3 is the same with case 1 but the distance of sensors reduced to 0.25m. Clustering error rates are represented in percentage for our proposed method (ICA) and normalization method with selecting two reference sensors respectively (Norm 1 and Norm 2).*

InputSIR	ICA	Norm 1	Norm 2
6.6163	8.2823	7.7444	8.1493

Tab. 2. *SIR improvement for each method (dB). ICA is our proposed method and Norm 1, 2 are the normalization method with changing the reference sensor.*

improvement for the two permutation ordering methods is shown in Tab. 2. Every permutation ordering method improve the SIR but normalization method has some oscillation in their performance as the referenced sensor is changed.

5. Conclusions

To exploit geometric information of basis vectors obtained from ICA and to order their permutation, we proposed a novel method based on ICA again. Our method finds the intrinsic directions of the basis vectors by regarding them as mixtures attacked by ICA again. Performance tests showed that our method is robuster than previous method, which uses sophisticated normalization and requires more prior knowledge. Our experimental results obviously verified the improvement of separation performance in terms of SIR, even though the amount of increase is not quite satisfiable. We expect our method can be an alternative to the previous framework that constrains the filter length in the highly reverberant environments.

Acknowledgment

This work is supported by ETRI.

References

1. P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” Neurocomputing, vol. 22, pp. 21-34, 1998.

2. L. Parra and C. Spence, “Convolutive blind source separation of nonstationary sources,” IEEE Trans. Speech Audio Processing, vol. 8, pp. 320-327, May 2000.

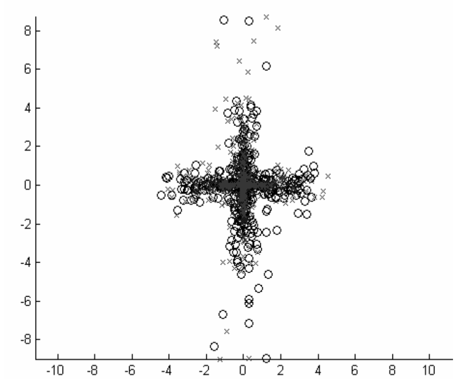
3. N Murata, S. Ikeda and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” Neurocomputing, vol. 41, no. 1-4, pp. 1-24, Oct, 2001

4. M. Z. Ikram and D. R. Morgan, “A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation,” in Proc. ICASSPP, May 2002, pp, 881-884

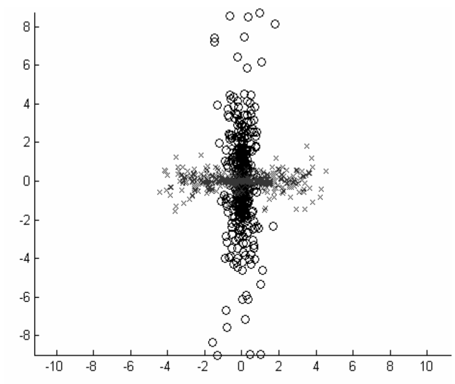
5. R. H. Lambert and A. J. Bell, “Blind separation of multiple speakers in a multipath environment”, in Proc. ICASSP, Apr. 1997, pp, 423-426.

6. M. Joho and P. Schniter, “Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient,” in Proc. ICA, Apr. 2003, pp 543-548.

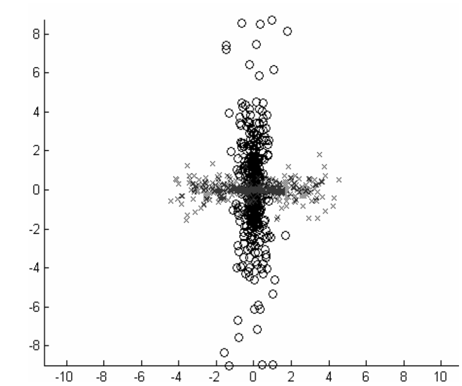
7. H. Sawada, S. Araki, R. Mukai, S. Makino, “Blind extraction of a dominant source signal from mixtures of many sources using ICA and time-frequency masking,” IEEE International Symposium on Circuits and Systems (ISCAS 2005), pp. 5882-5885, May 2005.



(a) Before ordering



(b) Normalization method



(c) Proposed method using IC

Fig. 5. *Ordering results of the artificially permuted frequency responses. Circle points are the first column vectors and cross points are the second ones. We regard that permutation is ordered if the first column vectors represent same directions and so do the second ones. (b) and (c) show the right results.*

8. E. Bingham and A. Hyvärinen, “A fast fixed-point algorithm for independent component analysis of complex-valued signals,” Neural Systems, 10(1):1-8, 20009.

9. A. Bell and T. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” Neural Computation, vol. 7, no. 6, pp. 1129-1159, 1995.

10. S. Amari, “Natural gradient works efficiently in learning,” Neural Computation , vol. 10, no. 2, pp. 251-276, 1998.