Blind Audio Source Separation Using Independent Component Analysis and Independent Vector Analysis

AAlyaa.A. M.Alrustim*[1], Ahmet Elbir*[1], Fethullah Karabiber*[1]

Abstract: Blind Source Separation (BSS) is one of the most important and challenging problem for the researchers in audio and speech processing area. In the literature, many different methods have been proposed to solve BSS problem. In this study, we have compared the performance of three popular BSS methods based on Independent Component Analysis (ICA) and Independent Vector Analysis Models, which are Fast-ICA, Kernel-ICA and Fast-IVA. We collected experimental data by recording speech from 13 people. Three different scenarios are proposed to compare the performance of BSS methods effectively. Experimental results show that the Fast-IVA has better performance than the ICA based methods according to performance metrics of Source-to-Artifact Ratio, Source-to-Distortion Ratio and Source-to-Noise Ratio. But ICA methods give better results than Fast-IVA according to the Source-to-Interference Ratio.

1. Introduction

Blind Source Separation (BSS) is one of the most important problems in speech processing area. The better description of this problem can be represented by this question: How can we accurately determine what a particular person talks among several speakers at the same time?. Figure 1 shows an illustration of BSS problem. This problem describes the situation of focusing on one speaker in case of several persons talking simultaneously in same room. To separate the mixed speech signals to obtain just a speech signal which belongs to a particular speaker is very challenging and complicated problem [1]. In the literature, many different methods based on signal processing and statistics were proposed to solve BSS problem.
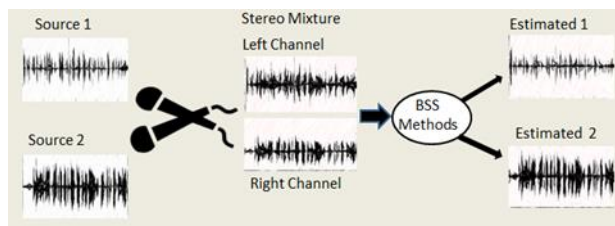

Figure 1. BSS Problem

BSS was firstly addressed by Herault and Jutten in 1985 [2]. In their work, the sound is directly transmitted to the microphones without any delay which known as standard blind source separation. Then Bell and Sejnowski in 1995 developed the Independent Component Analysis (ICA) method to solve BSS problem when the sources are mixed simultaneously [3]. Also some different algorithms based on ICA such as Fast-Fixed Point ICA [4], the Jade-ICA [5], the EGLD-ICA, the MS-ICA [6], and the Kernel-ICA [7] were proposed in literature.

As a result, BSS problem becomes more complicated for real room environment and this speech propagation problem is called convolutive blind source separation (CBSS) [8].

In the literature, some solutions were proposed in the time domain. Due to the complicated calculation caused by convolution, Parra et al [15] suggested another method based on frequency domain. In frequency domain, the convolution is replaced with multiplication to have low cost in terms of execution time. However, the frequency based methods still have scaling and permutation ambiguities. To prevent permutation problem, an advanced method named Independent Vector Analysis (IVA) was proposed by Kim et al [9].

In this study we have compared the performance of two ICA based algorithms and Fast-IVA according to performance measurement metrics commonly used for BSS problem. This paper is organized as follow: Section 2 contains the standard ICA method and its properties, ICA based algorithms (Fast-ICA and Kernel-ICA). Section 3 explains details of Fast Fixed Point IVA algorithm. Section 4 gives brief information about the commonly used performance measurement metrics. Experimental results obtained by using different proposed scenarios are represented in Section 5. Finally the conclusion is presented in the last section.

2. Independent Component Analysis

Independent Component Analysis (ICA) is one of the most popular BSS methods. ICA was used extensively for many applications in various field of science and engineering. ICA, which is a statistical computational method, was employed to find underlying hidden factors among set of random vectors. The main aim of ICA method is to obtain the independent components (ICs), which are linearly independent or as independent as possible, by finding a linear representation of non-Gaussian data.obtained.

2.1. ICA Model

Suppose that we have two persons talking simultaneously in a room and two microphones were placed in different places for recording their talking. In this way, two speech signals are recorded by these microphones. These signals can be represented

[1]*Department of Computer Engineering, Yildiz Technical University Davutpasa-Esenler, Istanbul TURKEY*
* Corresponding Author: Email: alyaaamahdi@gmail.com ,
aelbir@yildiz.edu.tr, fkara@yildiz.edu.tr

as $x_1(t)$ and $x_2(t)$ where (t) is the time index. The signals from the speakers can also be represented as $s_1(t)$, $s_2(t)$, so the linear combination of the speech signals can be expressed as:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \qquad (1)$$

Where s_1 and s_2 are the original speech signals, a_{11}; a_{12}; a_{21} and a_{22} denote the parameters that depend on the distance between the speakers and the microphones. Assume that M is the number of observed mixture signals and L is the number of independent source signals. Then the model will be as follow:

$$x_i = \sum_{i}^{L} a_{ij}s_j, i = 1 \dots \dots M \qquad (2)$$

Since instead of summations like in (2), using vector-matrix notation is more suitable to, the ICA model can be rewritten as:

$$x = As \qquad (3)$$

Where A is an unknown matrix called the mixing matrix, x is a vector of the observed signals and **s** is a vector of the source signals. The challenge in the ICA method is the estimation of both A and **s** by using only the observed random vector x. It is assumed that the unknown mixing matrix should be invertible or pseudo-invertible. After the matrix **A** estimation, its inverse denoted by W can be computed. As a result ICs denoted by y is obtained simply by using following equation.

$$y = W x \qquad (4)$$

ICA requires some assumptions related to the sources and the mixing process. These assumptions make this method different from the other source separation approaches [10]. The first assumption is that the sources being considered are statistically independent. The other one is that the sources must have non-Gaussian distribution. The last one is to have invertible mixing matrix [10]. Two inherent ambiguities are hold in ICA framework. One of them is magnitude and scaling ambiguity and the other is the permutation ambiguity [10]. Some preprocessing steps can be performed to improve the performance of ICA based methods. Some useful preprocessing techniques are Centering and Whitening [10]. There are several kinds of algorithm based on ICA. In this study we have performed two of them, which are widely used.

2.2. Fast Fixed-Point ICA

The Fast Fixed-Point ICA (Fast-ICA) algorithm based on mutual information utilizes higher order statistics for the retrieval of independent sources [4]. Fast ICA has two estimation approaches:
((-i)deflation approach to estimate ICs one by one) and
((ii)symmetric approach to estimate ICs simultaneously)[4].
Fast-ICA uses fixed-point algorithm of ICA as follow:

Algorithm Fast fixed-point ICA [4]

1: Take a random initial vector w(0) of norm 1, Consider $k = 1$:
2: Let $w(k) = E_X(w(k-1)^T x)^3 3w(k-1)$, Estimate the expectation using a large sample of x vectors.
3: Divide w(k) by its norm.
4: if $|w(k)^T w(k-1)|$ is not close enough to 1, consider $k = k + 1$ and go back to step 2 , otherwise return the vector $w(k)$ as output.

2.3. Kernel Independent Component Analysis

The Kernel Independent Component Analysis (Kernel- ICA) is a different version of ICA model that based on the minimization of a contrast function based on kernel ideas. Kernel-ICA rely on an entire function space of candidate nonlinearities. In particular, Kernel-ICA works with the functions in a reproducing kernel Hilbert space using canonical correlation based functions. The Kernel trick is used to make the search over this space efficiently. Some new modifications were proposed to make the algorithm more robust and efficient to different source distributions [7]. In short, Kernel-ICA use optimization methods for canonical correlations to reproduce kernel Hilbert space.

3. Independent Vector Analysis

Independent Vector Analysis (IVA), which is one of the most advanced method, show better performance in the field of BSS. Theoretically this method overcomes the permutation problem that inherent from ICA. It is designed to remove the dependency between different source vectors while keeping the dependency within individual source vectors. The problem is transformed into frequency domain to reduce computational complexity of the time domain. The noise free model in the frequency domain can be defined as follow:

$$X(K) = H(K)s(K) \qquad (5)$$
$$\hat{S}(K) = W(K)x(K) \qquad (6)$$

Where the index $k = 1, 2,, K$ indicate the k-th frequency bin, and K is the number of frequency bins, $x(K) = [x_1(k); x_2(k); x_m(k)]^T$ represent the observed signal vector in the frequency domain, $s(K) = [s_1(k); s_2(k); s_m(k)]^T$ denotes the original source vector in the frequency domain and $\hat{S}(K) = [\hat{S}_1(k); \hat{S}_2(k); \hat{S}_m(k)]^T$ represents the estimated source vector in the frequency domain. And $(.)^T$ is the vector transpose. The number of microphones and the number of sources are represented by m and n, respectively. H (k), which is a m n dimension matrix, denotes the mixing matrix, and W(k), which is a n m dimension matrix, denotes the unmixing matrix.

In this study, we assume that the number of sources and the number of microphones should be the same. The main goal is that the sources should be estimated by using only the observed (mixed) signals.

An objective function is defined to separate multivariate sources from multivariate observations. Kullback-Leibler divergence between two functions as the measure of dependence is employed in IVA. These two functions are the joint probability density function $p = \hat{S}_1, \dots \hat{S}_n$ and the product of probability density functions of the individual source vectors $\Pi q(\hat{S}_i)$. This function can be defined as follow:

$$j = KL\left(p(\hat{S}_1, \dots \hat{S}_n) \| \Pi q(\hat{S}_i)\right)$$
$$= \text{cost} - \sum_{k=1}^{K} \log |\det(W)| - \sum_{n=1}^{n} E[log q(\hat{S}_i)] \qquad (7)$$

We can keep the dependency between the components of each vector, and remove the dependency between the source vectors if the cost function is minimized [11]. In literature, there are different version of IVA such as NG-IVA, Fast-IVA and Aux-IVA [11]. In this study Fast-IVA algorithm is employed for BSS.

3.1. Fast Fixed-Point Independent Vector Analysis

This algorithm utilizes Newton's method to update the original IVA method, which converges quadratically and select an efficient learning rate. In order to apply Newton's method in the

---

update rules, polynomial approximation of a quadratic Taylor series is produced in the notations of complex variables. In this way, it can be used for a contrast function of complex-valued variables [11]. The contrast function used by Fast IVA is as follows:

$$J = \sum_{i=1}^{n} (E[G \sum_{k=1}^{K} |\hat{s}_i(k)|^2] \sum_{k=1}^{K} \lambda\, i^{(k)} (w_i(k) w_i(k-1)) \quad (8)$$

Where, λ_i is the ith Lagrange multiplier, and w(i) denotes the ith row of the unmixing matrix W, G (.) is the nonlinearity function, which can take on several different forms as discussed in [11]. The learning rule can be defined as follow with normalization:

$$(W_i^{(k)})^H \leftarrow E\left[G\left(\sum_K |\hat{s}_i^{(K)}|^2\right) + \sum_K \hat{s}_i^{(K)}|^2\, G''\left(\sum_K |\hat{s}_i^{(K)}|^2\right)\right]$$

$$x\,(W_i^{(k)})^H - E\left[\left(\hat{s}_i^{(K)}\right) * G'\left(\sum_K |\hat{s}_i^{(K)}|^2\right) x^K\right] \quad (9)$$

where $G'(\,)$ and $G''(\,)$ represent the first and second derivative of G( ), respectively. And if we use that for all sources, we can construct an unmixing matrix W(k) to be decorrelated with

$$W^K \leftarrow (W^{(K)}(W^{(K)})^T)^{-\frac{1}{2}} W^{(K)} \quad (10)$$

4. Performance Measurement

There are several performance measurement metrics to evaluate the quality of estimated signals obtained by BSS methods. The performance of BSS algorithms can be measured by comparing each estimated source ŝj to a given true source sj. The measuring processing includes two successive steps [12].
The first step involves decompose ŝj as:

$$\hat{s}_j = s_{tatget} + e_{interf} + e_{noise} + e_{artif} \quad (11)$$

Where $s_{tatget} = f(s_j)$ denotes the version of s$_j$ modified by an allowed distortion, and the interferences, noise, and artifacts error terms are represented by s$_{interf}$, s$_{noise}$ and s$_{artif}$ respectively.
The second step involves computing the energy ratios in order to estimate the relative amount of each of these four terms either on the local frames of the signal or the whole signal duration. The way of decomposing into four terms are given in [12] in detail. Relevant energy ratios between these terms are defined.
 After the decomposition of ŝ$_j$ following the procedures given in [21]. Numerical performance criteria was defined by computing energy ratios expressed in decibels. Definition of source-to-distortion ratio (SDR), the source-to-interference ratio (SIR), source-to- artifact ratio (SAR) and source-to-noise ratio (SNR) are given below, respectively

$$SDR = 10\, log_{10}\, \frac{||s_{target}||^2}{||e_{interf} + e_{noise} + e_{artif}||^2} \quad (12)$$

$$SIR = 10 log_{10}\, \frac{||s_{target}||^2}{||e_{interf}||^2} \quad (13)$$

$$SAR = 10 log_{10}\, \frac{||s_{target} + e_{interf} + e_{noise}||^2}{||e_{artif}||^2} \quad (14)$$

$$SNR = 10 log_{10}\, \frac{||s_{target} + e_{interf}||^2}{||e_{noise}||^2} \quad (15)$$

5. Experimental Results

The performance of Fast-ICA, Kernel-ICA and Fast-IVA methods for separating mixing speech signals was compared. We collected 13 speech signals recorded in real room, each record long 10sec in Arabic language with 16 kHz. These records are mixed by using random parameters. In our experiment, three different scenarios are proposed to compare performance more effectively. The first scenario includes measuring and comparing the performance of Fast-ICA, Kernel-ICA and Fast-IVA for separating mixing speech signals without noise, as shown in Figure 2. Figure 3 illustrates the second scenario which shows the performance of these methods for separating mixed speech signals with Gaussian noise added to signals before mixing. In the third scenario, we add Gaussian noise to the signals after mixing as shown in Figure 4. Since Gaussian noise is added to the sources or mixtures in the second and third scenarios, Savitzky-Golay smoothing filter [13] is performed to enhance the signals before the separation.
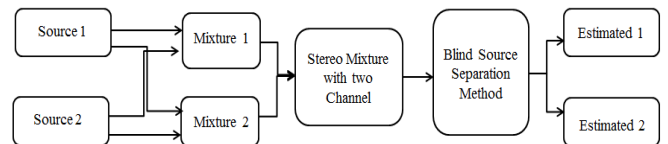

Figure 2. Illustration of Scenario 1 for the experiment

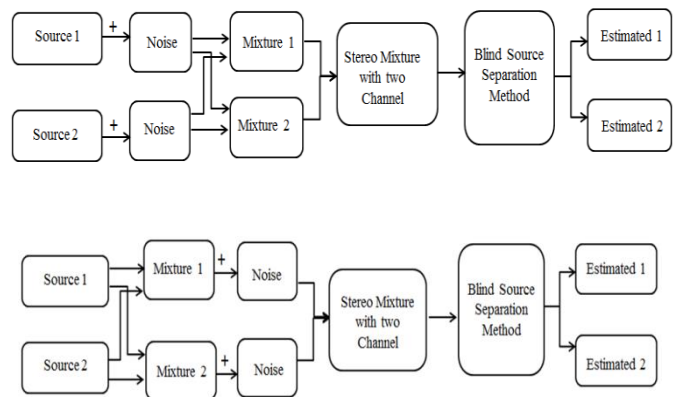**Figure 3.** Illustration of Scenario 2 for the experiment

Figure 4. Illustration of Scenario 3 for the experiment

In each scenario, we have created 77 different mixture signals from 13 original sources using different mixture parameters. The Fast-IVA use the Short time Fourier Transform of length 1024, and the sampling frequencies were 16 kHz. The results of our experiment for each scenario show that: the Fast-IVA has better performance comparing with Fast-ICA and Kernel-ICA according to SAR, SDR and SNR performance metrics. But we notice that Fast-ICA and Kernel-ICA methods have better performance than Fast-IVA according SIR as shown in Table 1,

Table 2 and Table3.

In this study, we have implemented Fast-ICA algorithm in MATLAB. A different version of Fast-ICA algorithm can be downloaded from http://research.ics.aalto.fi/ica/fastica/code/dlcode.shtml. For the other BSS methods (Kernel-Ica and Fast-IVA), we have used the shared codes. The kernel-ICA package is Copyright (c) 2002 by Francis Bach [14], and Fast-IVA code by Taesu Kim, recently revised at Nov. 2, 2005 [9]. The experiments were performed in Matlab ver. 8.1.0.604 (R2013a). The algorithms are also compared in terms of execution time. The time that needed for running the algorithms was also given in Table 4.

Table 1. Experimental result for the first scenario

Algorithm	Criteria	SAR	SDR	SIR	SNR
Fast-ICA	Average	6.114	6.087	29.671	-13.47
	STdev	1.084	1.087	2.3746	3.0666
Kernel-ICA	Average	6.1374	6.1130	29.9469	-13.328
	STdev	1.0889	1.0904	2.03176	3.14367
Fast-IVA	Average	21.651	18.873	24.181	8.6396
	STdev	7.6494	8.278	9.4262	8.7521

Table 2. Experimental result for the second scenario

Algorithm	Criteria	SAR	SDR	SIR	SNR
Fast-ICA	Average	5.0701	5.0289	27.705	-13.660
	STdev	1.8008	1.7975	3.0743	2.8805
Kernel-ICA	Average	4.792	4.822	28.657	-13.611
	STdev	2.678	2.976	4.987	2.954
Fast- IVA	Average	8.8899	6.4275	15.2160	2.0444
	STdev	3.2868	5.5514	10.2810	1.8588

Table 3. Experimental result for the Third scenario

Algorithm	Criteria	SAR	SDR	SIR	SNR
Fast-ICA	Average	5.0037	4.9645	27.8722	-13.745
	STdev	1.8200	1.8168	3.1256	2.7637
Kernel-ICA	Average	4.9694	4.9435	28.9186	-14.035
	STdev	1.8174	1.8199	2.61689	2.6280
Fast-IVA	Average	8.8166	6.3595	15.0365	2.0093
	STdev	3.3440	5.4436	9.9977	1.9585

Table 4. Execution time for the algorithm

Algorithm	Running time
Fast-ICA	24.788 s
Fast -IVA	30.006 s
Kernel-ICA	180.567 s

6. Conclusion

Blind source separation (BSS) is one of the most important and challenging problem for the researchers in audio and speech processing fields. In this study, we have implemented and compared three popular BSS methods, which are Fast-ICA, Kernel-ICA and Fast-IVA. Three different scenarios were proposed to measure the performance of BSS methods extensively using four different commonly used performance metrics. According to experimental results Fast-IVA shows better performance than the other methods according to SDR, SAR and SNR metrics. But Fast-ICA and Kernel-ICA methods have high performance than Fast-IVA according to SIR metric. In future,

some advanced BSS algorithms will be implemented and compared using different scenarios.

References

[1] S Haykin, Z Chen, The cocktail party problem" Neural Comput. 17,18751902 (2005).

[2] J Herault, Jutten C, B Ans, Detection of primitive magnitudes in a Message composite by a neural computing architecture unsupervised learning, in Proc. GRETSI Vol 2 (Nice, France, 1985), pp. 1017-1022

[3] Bell AJ, Sejnowski TJ, An information maximisation approach to blind separation and blind deconvolution. Neural Computation 7:11291159(1995).

[4] Aapo Hyvarien and Erkki Oja laboratory fast fixed point independent compenent analysis this paper appear in neural computation 9 :14831492; 1997.

[5] Sargam Parmar, Bhuvan Unhelkar Performance Comparisions of ICA Algorithms to DS-CDMA Detection journal of telecommunications, volume 1, issue 1,february 2010.

[6] HongLi, Yunlian Sun, The Study and Test of ICA Algorithms 2005 IEEE.

[7] Francis R. Bach, Michael I. Jordan, Kernel Independent Component Analysis. Journal of Machine Learning Research 3 1-48. (2002).

[8] MS Pedersen, J Larsen, U Kjems, LC Parra. A survey of convolutive blind source separation methods, in Springer Handbook on Speech Processing and Speech Communication (Springer, New York, 2007),pp. 134.

[9] T Kim, I Lee, TW Lee. Independent vector analysis: definition and algorithms, in Fortieth Asilomar Conference on Signals, Systems andComputers 2006 (Asilomar, USA, 2006), pp. 13931396.

[10] Ganesh R. Naik and Dinesh K Kumar, An Overview of Independent Component Analysis and Its Applications , Informatica 35 (2011) 6381.

[11] Yan feng Liang n, Jack Harris, Syed Mohsen Naqvi, Gaojie Chen,Jonathon A. Chambers , Independent vector analysis with a generalized multivariate Gaussian source prior for frequency domain blind source separation, at Science Direct Signal Processing 28May2014.

[12] Emmanuel Vincent, Rmi Gribonval, and Cdric Fvotte, Performance Measurement in Blind Audio Source Separation, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 14, No. 4, July 2006.

[13] Manfred U.A. Bromba and Horst Ziegler, Application hint for Savitsky-Golay digital smoothing filters, Anal. Chem. 53, 15831586 (1981).

[14] Francis R. Bach, Michael I. Jordan (2001). Kernel Independent Component Analysis, Technical Report UCB//CSD-01-1166, University of California, Berkeley.

[15] L Parra, C Spence, Convolutive blind separation of non-stationary sources, IEEE Trans. Speech Audio Process. 8, 320327 (2000).