*SindhUniv. Res. Jour. (Sci. Ser.) Vol.46 (3) 283-286 (2014)*

**SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)**

# A Novel Method for Blind Segregation of Speech and Image Data Using Independent Vector Analysis.

M. SHAH, T. JAN[++], A. JEHANGIR, A. ALI*

Department of Electrical Engineering, University of Engineering and Technology, Peshawar

Received 09th April 2014 and Revised 13th July 2014

**Abstract:** Independent Vector Analysis (IVA) is a handy algorithm that is used to separate a Convolutive mixture into its constituent signals that are most common in acoustic surrounding. While postulating that sources are independent and mixing is linear. IVA strongly excludes the permutation issue in the frequency domain by the adoption of spherical dependency model to the entire frequency bins. Which was one of the dilemma with Independent Component Analysis (ICA) even though it was handled by several engineering solutions. Till now IVA algorithm segregates the mixture into its components that are generated by homogenous sources, while in this paper heterogeneous sources are introduced. Quality of splitting is determined utilizing signal-to-noise ratio (SNR).

**Keywords***:* Independent Component Analysis (ICA), Independent Vector Analysis (IVA), Heterogeneous mixtures, Blind Source Separation (BSS).

## 1. INTRODUCTION

Over the past couple of years blind source separation (BSS) has been appealing for researchers in the various fields of engineering and sciences. It is the decomposition of set of signals from a set of combined signals, without knowing the sources and procedure of mixing. It is viable under certain assumptions and with some limitations in the form of scaling and permutations (Tong *et al.*, 1991). There are applications such as EMGs signals, where these limitations cannot be tolerated. While in most of the other scenarios these can be neglected.

Independent Component Analysis (ICA) and Independent Vector Analysis (IVA) are the two major BSS Algorithms used for the extraction of signals from a mixture. ICA achieves the task up to some extent provided that the sources are statistically independent to each other and the dataset of each source are dependent (Comon *et al.*, 1994) (Hyv¨arinen *et al.*, 2001). Anyhow presume that sources are random, means they are multi-dimensional. Inherently Scaling and Permutations were the two major drawbacks in ICA which later on were removed by several engineering solutions such as technique that limits the filter length in the time domain (Parra *et al.*, 2000), uses direction of arrival estimation (Ikram *et al.*, 2002), and uses inter-frequency correlation (Murata *et al.*, 2001).

While IVA is another method for resolving BSS scenario. This Algorithm has screened out the permutation ambiguity inherently but scaling problem is still here like ICA. Sources are supposed to be independent but their elements are dependent and co-related to each other because each source is taken as a vector as in ICA. Furthermore different models are utilized for the mixing of these independent vectors in different dimensions (Kim *et al.*, 2006).

So far in the literature related to signals segregation, different homogeneous mixtures have been considered but in real, heterogeneous mixtures (e.g. image and sound) can be formed and such a challenging scenario has been tackled in this work using IVA algorithm.

This paper is organized as follows, In Section 2 the Independent Vector Analysis is introduced in combination with proposed method, followed by experimental results in Section 3. At the end section 4 comprises of conclusion of the paper.

## 2. MATERIAL AND METHODS

### INDEPENDENT VECTOR ANALYSIS

### 2.1 Introduction to IVA.

Given mixtures $\boldsymbol{x}_i$ (Kim *et al.*, 2006),

$$\boldsymbol{x}_i = \sum_j^L \boldsymbol{h}_{ij} \, o \, \boldsymbol{s}_j \qquad \text{(i)}$$

Determining the source vectors $\boldsymbol{s}_j$ by (Ikram*et al.*, 2002)

$$\boldsymbol{s}_i \approx \boldsymbol{y}_i = \sum_j^M \boldsymbol{w}_{ij} \, o \, \boldsymbol{x}_j \qquad \text{(ii)}$$

In the above expression, ○ is element wise product, while $L$ represents the number of sources, and $M$ denotes the number of mixtures.

**Assumptions:**

1. There is mutually independency between the Elements of one source vector with the elements of the other source vectors.
2. The elements of a source are dependent on each other.

[++]Corresponding Author: tariqullahjan@nwfpuet.edu.pk, Ph. +92 91 9216498
*PTCL, Khyber Exchange, Peshawar, Pakistan

3. The number of observations is greater than or equal to the number of sources.

$$\boldsymbol{x}^{(d)} = H^{(d)}.\boldsymbol{s}^{(d)} \quad \text{(iii)}$$

Where $\boldsymbol{x}_i = [x_i^{(1)}, \dots, x_i^{(D)}]^T$,$\boldsymbol{h}_{ij} = [h_{ij}^{(1)}, \dots, h_{ij}^{(D)}]^T$, $\boldsymbol{s}_i = [s_i^{(1)}, \dots, s_i^{(D)}]^T$, $h_{ij}^{(d)}$ is the *i*th row , *j*th column component of the *d*th mixing matrix $H^{(d)}$ and *d* represents the frequency bin.

In order to get segregated multivariate elements from multivariate mixture, contrast function is defined for multivariate random variables.

This algorithm follows the Kullback-Leibler divergence such that to examine the dependency between two functions, in which Exact joint probability density is the first function while the second function is non-linear and product of approximated marginal probability density function (Kim *et al.,* 2006).

$$Q = KL(p(\boldsymbol{s}_1, \dots, \boldsymbol{s}_L)||\prod_i q(\boldsymbol{s}_i)$$
$$= const. + \sum_d \log|detH^{(d)}| - \sum_i E_{\boldsymbol{s}_i} log\ q(\boldsymbol{s}_i) \text{(iv)}$$

Where$E[.]$ represents the statistical expectation operator, det(.) shows the determinant of matrix. And $p(\boldsymbol{s}_1, \dots, \boldsymbol{s}_L)$ is the exact joint pdf and $\prod_i^L q(\boldsymbol{s}_i)$ is the product of marginal pdf of each source vectors. By reducing the cost function, the dependency among all source vectors should be eliminate provided the interrelationship between the elements of each vectors can be sustained (Liang *et al.,* 2013).

For clarity, it may be supposed that observation $\boldsymbol{x}_i$ has zero-mean and the process $\boldsymbol{y}^{(d)} = W^{(d)}\boldsymbol{x}^{(d)}$has zero-mean and structurally. It is suitable for simplicity and more appropriate convergence. This can be achieved by subtracting the mean of$\boldsymbol{x}_i$ and pre-whitening $\boldsymbol{y}^{(d)}$ in each and every dimension. To sustain whitening in the learning process, we should compel that the rows of mixing matrix or the un-mixing matrix in each dimension to be orthogonal (Kim *et al.,* 2006).Therefore, the contrast function becomes

$$Q = const. - \sum_i E_{\boldsymbol{s}_i} \log\ q(\boldsymbol{s}_i) \quad \text{(v)}$$

In every dimension the pre-whitening is not useful every time for IVA algorithms as a result there is un-correlation between the components of source vector. Anyhow it's actually useful in some applications like BSS in the frequency domain signals, because there is almost un-correlation in the Fourier coefficients of a natural signal, but they are largely dependent (Kim *et al.,* 2006).

Observe that the contrast functions' random variables are multivariable. The fascinating thing about contrast functions is that every source is multivariate, but by removing the dependency between the sources vectors, the multi-variation would get reduced. Along with that the dependency between the elements of each source should not be disturbed. Hence the inherent dependency within each source vector is secured by contrast function, only it expels the dependency between the sources (Kim *et al.,* 2006).

Now in order to determine the learning algorithm, the contrast function is minimized by implementing the gradient descent method. Learning rule can be easily found out by differentiating the contrast function Q with respect to un-mixing matrices' $w_{ij}^{(d)}$coefficient (Kim *et al.,* 2006).

$$\Delta w_{ij}^{(d)} = -\frac{\partial Q}{\partial w_{ij}^{(d)}}$$
$$= h_{ij}^{(d)} - E\varphi^{(d)}(\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(D)})\boldsymbol{x}_j^{(d)} \quad \text{(vi)}$$

The natural gradient learning rule (Amari *et al.,* 1996), is notable like fast convergence method can be obtained by the multiplication of scaling matrices, $W^{(d)^T}W^{(d)}$(Kim *et al.,* 2006).

$$\Delta w_{ij}^{(d)} = \sum_{l=1}^{L}(I_{il} - E\varphi^{(d)}(\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(D)})\boldsymbol{y}_l^{(d)})w_{lj}^{(d)} \text{(vii)}$$

When $i = l$, then $I_{il}$ is unity, else it is zero, a multivariate score function follows as (Kim *et al.,* 2006).

$$\varphi^{(d)}\left(\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(D)}\right) = -\frac{\partial \log q(\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(D)})}{\partial \boldsymbol{y}_i^{(d)}} \quad \text{(viii)}$$

The learning rule becomes simple under the impulsion of whiteness (Kim *et al.,* 2006).

$$\Delta w_{ij}^{(d)} = -E\varphi^{(d)}(\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(D)})\boldsymbol{x}_j^{(d)} \quad \text{(ix)}$$

Now we follow Newton method with fixed point iteration along with gradient methods. It can skip adopting legitimate learning rate and it has got the advantage of fast convergence speed, contrast to other gradient methods. Now we develop it for IVA model and also find its contrast function (Kim *et al.,* 2006).

While whiteness is imposed, the contrast function (v) is given as with Lagrangian multiplier $\beta$.

$$Q \equiv -\sum_i E_{\boldsymbol{s}_i} \log q(\boldsymbol{s}_i) - \sum_d \beta(W^{(d)^T}W^{(d)} - I) \quad \text{(x)}$$

After the application of Newton method we get the following easy learning rule thereby the contrast function of the Hessian matrix which is diagonal under the whiteness impulsion (Kim *et al.,* 2006).

$$w_i^{(d)} \leftarrow w_i^{(d)} - \frac{E\left[\varphi^{(d)}\left(y_i^{(1)}, \dots, y_i^{(D)}\right)\boldsymbol{x}^{(d)}\right] + \beta \boldsymbol{w}_i^{(d)}}{E\left[\varphi^{(d)'}(y_i^{(1)}, \dots, y_i^{(D)})\right] + \beta} \quad \text{(xi)}$$

The local minimum point of the contrast is the equilibrium of (xi) so the un-mixing matrix cannot be

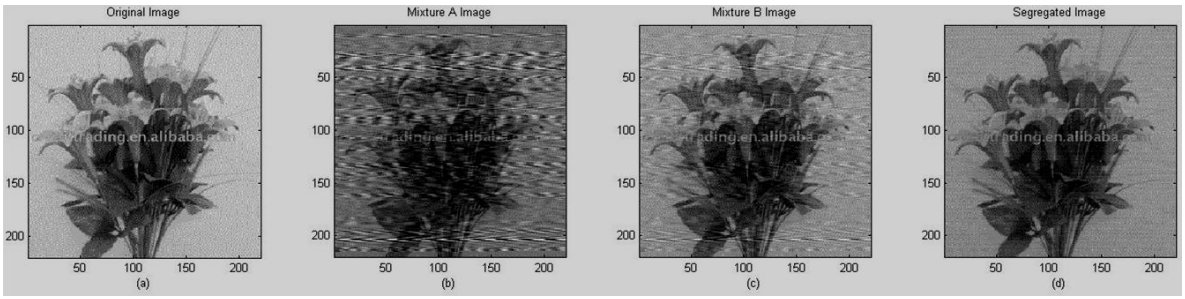


**Fig.1. (a) is the original image, (b) and (c) are mixtures (mixture A, mixture B) in 2-D (image) form and (d) represents segregated image.**

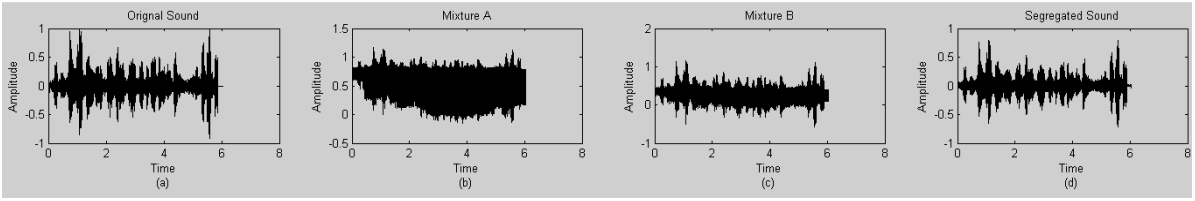


**Fig.2. (a) is the original sound, (b) and (c) are mixtures (mixture A, mixture B) in 1-D (audio) form and (d) represents segregated-sound.**

updated further. As the equality is satisfied the equilibrium point is achieved. So we make both sides of (xi) be the same. Lagrangian multiplier can be removed by the multiplication of numerator $E[\varphi^{(d)'}(y_i^{(1)},\dots,y_i^{(D)})]+\beta$ to left and right side of the equation. Then the fixed point iteration algorithm will be (Kim *et al.*, 2006).

$$w_i^{(d)} \leftarrow E[\varphi^{(d)'}\left(y_i^{(1)},\dots,y_i^{(D)}\right)]\boldsymbol{w}_i^{(d)} - E[\varphi^{(d)}(y_i^{(1)},\dots,y_i^{(D)})\boldsymbol{x}^{(d)}] \quad \text{(xii)}$$

$$\varphi^{(d)'}\left(\boldsymbol{y}_i^{(1)},\dots,\boldsymbol{y}_i^{(D)}\right) = -\frac{\partial\varphi^{(d)}(\boldsymbol{y}_i^{(1)},\dots,\boldsymbol{y}_i^{(D)})}{\partial y_i^{(d)}} \quad \text{(xiii)}$$

One possible example of the nonlinear function is (Kim *et al.*, 2006).

$$\varphi^{(d)}\left(y_i^{(1)},\dots,y_i^{(D)}\right) = \frac{y_i^{(d)}}{\sqrt{\sum_d |y_i^{(d)}|^2}} \quad \text{(xiv)}$$

Possibility may be here to use other forms or even more flexible form of nonlinear function but we offer a particular form of the nonlinear function. Because different nonlinear functions are generated by different vector density model (Kim *et al.*, 2006).

### 2.2 Proposed method.

In current work a method has been proposed which is based on IVA algorithm for heterogeneous mixtures (that of combination of image and sound data). In literature mostly homogenous mixtures have been found where ICA and IVA algorithms have been used for segregation but in current work IVA algorithm has been applied for the first time on heterogeneous mixtures and successful results have been achieved which will be discussed in the next section.

There are two types of the mixtures that could be used under the umbrella of BSS algorithms, instantaneous and Convolutive mixtures. In instantaneous procedure signals are mixed directly without considering the effect of reflections while on the other hand in Convolutive mixture the effect of reflection has also been included.

In the current work instantaneous heterogeneous mixtures have been used.

## 3. DISCUSSION AND RESULTS

In current work, a setup has been formed in which heterogeneous mixtures have been generated (i.e. speech and image), and the IVA algorithm has been applied for the segregation. The performance of the experiments that has been carried out on heterogonous mixtures is evaluated by simulation. Here an image and an audio signal are the original sources as shown in Fig. 1(a) & Fig.2 (a) respectively. Size of image is 220×220 Pixels, black & white, and dimension is converted from 2-D to 1-D during experiments. This modified 1-D matrix is consist of 48400 pixels of the image. After the selection of sources mixing is accomplished by instantaneous procedure. In Fig. 1 both mixtures are shown in image form same mixture are presented in audio (1-D) form in Fig. 2.

The sampling rate is 8000 Hz, a 512 point FFT and handing window for Short Time Fourier Transform is used. The window length is 512 samples and the shifting size was 128 samples. After segregation of

mixture using IVA algorithm, 1-D array of image is converted to 2-D matrix to form image shown in Fig. 1.

For evaluation, SNR is used. Overall Δ*SNR* is

$$\Delta \boldsymbol{SNR} = \boldsymbol{SNR_{out}} - \boldsymbol{SNR_{in}} \quad \text{(xv)}$$

*SNRin* and *SNRout* (Jan *et al.*, 2011) in time domain are

$$\boldsymbol{SNR_{in}} = \boldsymbol{10log_{10}} \frac{\sum_t (a_i[t])^2}{\sum_t (a_i[t] - x_i[t])^2} \quad \text{(xvi)}$$

$$\boldsymbol{SNR_{out}} = \boldsymbol{10log_{10}} \frac{\sum_t (x_i[t])^2}{\sum_t (x_i[t] - y_i[t])^2} \quad \text{(xvii)}$$

Where $a_i[t]$ are the original source components in time domain and Δ*SNR* is improvement in SNR. Here 8 sound signals are randomly selected from limit database. Similarly 8 images are downloaded randomly from a web source named as http://www.chineserose.en.alibaba.com/. Then 50 tests were performed in which these sounds and images were randomly selected, mixed with each other instantaneously and then segregated into its source signals using IVA algorithm.

After random selection of image and audio two mixtures are generated by instantaneous mixing procedure. These two mixtures (mixture A, mixture B) are presented in 2-D and in 1-D as well. Fig.1 (b) and Fig.1 (c) are 2-D while Fig.2 (b) and Fig.2 (c) are 1-D representation.

As shown in Fig.1 (d) is reconstructed image, we can clearly see that Fig.1 (d) is much closer to Fig.1 (a). Similarly in Fig.2, Fig.2 (d) is the reconstructed sound while Fig.2 (a) is the original sound, we can clearly see that after separation, Fig.2 (d) has similarity with Fig.2 (a). This clearly shows that in the proposed method, the reconstructed image and sound are much closer to the original image and sound.

After evaluation almost 15dB improvement is achieved in Δ*SNR* for the output signals. While in results, SNR of image and sound has improved, getting a clear image and sound. On the other hand contrast of reconstructed image and the amplitude of the reconstructed sound are found altered from the original image and sound. This occurred due to scaling issue of IVA algorithm.

## 4. CONCLUSION

Till now homogenous mixtures have been extensively used for segregation using IVA technique, but in recent work focus is on heterogeneous mixtures. Here sound and image have been mixed and IVA algorithm has been used for the segregation of such signals, random experiments have been performed showing that improved performance has been achieved in terms of reconstruction of sound and image data. Although these two signals are different from one another statistically but the proposed method still shows better results.

In current scenario instantaneous heterogeneous mixtures are segregated in to its source signals, while in future convolutive heterogeneous mixtures will be considered.

## REFERENCES:

Amari. S., A. Cichocki, and H. H. Yang. (1996) A new learning algorithm for blind signal separation, Advances Neural information Processing Systems, vol. (8): 757-763.

Comon, P. (1994) Independent component analysis, A new concept?, Signal Processing, vol. (36): 287–314.

Hyv¨arinen, A. and E. Oja, (2002) Independent Component Analysis, John Wiley and Sons, New York, USA.

Ikram, M. Z. and D. R. Morgan. (2002) A Beamforming Approach to Permutation Alignment for Multichannel Frequency-Domain Blind Speech Separation, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, USA, 881–884.

Jan, T, and W. Wang. (2011) Empirical mode decomposition for joint denoising and dereverberation, 19th European Signal Processing Conference (EUSIPCO 2011), Barcelona, Spain, 206-210.

Kim, T., I. Lee, and T. W. Lee. (2006) Independent vector analysis: definition and algorithms, in Fortieth Asilomar Conference on Signals, Systems and Computers 2006, Asilomar, USA, 1393–1396.

Liang, Y., S. M. Naqvi, and J. A. Chambers. (2013) Independent vector analysis with a multivariate generalized Gaussian source prior for frequency domain blind source separation, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6088-6092.

Murata, N., S. Ikeda, and A. Ziehe. (2001) An approach to blind source separation based on temporal structure of speech signals. Neurocomputing, vol. (41):1–24.

Parra, L. and C. Spence. (2000) Convolutive blind separation of non-stationary sources, IEEE Trans. Speech Audio Process., vol. (8), No 3: 320–327.

Tong, L., R. Liu, V. C. Soon, and Y. F. Huang. (1991)Indeterminacy and Identifiability of Blind Identification, IEEE Tran. Circuits and Systems, vol. (38), No. 5, 499–509.