EUSIPCO 2013 1569746061

AUXILIARY FUNCTION BASED IVA USING A SOURCE PRIOR EXPLOITING FOURTH ORDER RELATIONSHIPS

Yanfeng Liang[*], *Jack Harris*[*,†], *Gaojie Chen*[*], *Syed Mohsen Naqvi*[*], *Christian Jutten*[†], *Jonathon Chambers*[*]

[*]School of Electronic, Electrical and System Engineering
Loughborough University, UK
{y.liang2, g.chen, s.m.r.naqvi, j.a.chambers}@lboro.ac.uk
[†]GIPSA-Lab, CNRS UMR5216,
Université de Grenoble, France
{jack.harris, christian.jutten}@gipsa-lab.grenoble-inp.fr

ABSTRACT

Independent vector analysis (IVA) can theoretically avoid the permutation ambiguity present in frequency domain independent component analysis by using a multivariate source prior to retain the dependency between different frequency bins of each source. The auxiliary function based independent vector analysis (AuxIVA) is a stable and fast update IVA algorithm which includes no tuning parameters. In this paper, a particular multivariate generalized Gaussian distribution source prior is therefore adopted to derive the AuxIVA algorithm which can exploit fourth order relationships to better preserve the dependency between different frequency bins of speech signals. Experimental results confirm the improved separation performance achieved by using the proposed algorithm.

Index Terms— AuxIVA, multivariate generalized Gaussian distribution, fourth order relationships

1. INTRODUCTION

Independent component analysis (ICA) is the central tool for the blind source separation (BSS) problem [1][2]. The most famous BSS problem is the cocktail party problem [3][4], in which one speaker must be selected from a mixture of sounds. In the real room environment, due to reverberation, it becomes a convolutive blind source separation (CBSS) problem. Therefore time domain methods are not appropriate because of the computational complexity [1]. Thus frequency domain methods are proposed to solve the CBSS problem [5]. However, the permutation problem inherent to BSS needs to be solved to achieve the separation result. Many methods exploiting the positions or spectral structures of the sources have been proposed to solve the permutation problem, but all of these methods need post processing to address the problem [6].

Independent vector analysis (IVA) is proposed to solve the frequency domain blind source separation (FD-BSS) problem algorithmically. The permutation ambiguity can theoretically be avoided by using a multivariate source prior to retain the dependency between different frequency bins of each source [7][8][9]. IVA can thereby solve the permutation problem during the convergence process without post processing. In recent years, different modified IVA methods have been proposed. An adaptive step size IVA method is proposed to increase the convergence speed [10]. The fast fix-point IVA method adopts Newton's update method to obtain a fast convergence form of IVA [11]. Video information is introduced to form the audio-video based IVA method [12]. In 2011, the auxiliary function based form of independent vector analysis (AuxIVA) was proposed, which is a stable and fast form of IVA algorithm. By using the auxiliary function technique AuxIVA can guarantee a monotonic decrease of the cost function without the need for tuning parameters such as step size [13].

The source prior is important to all IVA methods, because it is used to derive the nonlinear score function which is used to keep the dependency between different frequency bins. For the original IVA algorithm, the multivariate Laplace distribution is adopted as the source prior [8]. However, it is not the only form for source prior; an improved source prior which can better preserve the relationships between different frequency bins is still needed. In this paper, we adopt a generalized multivariate Gaussian distribution as the source prior, which can preserve the fourth order terms within the score function. Then the AuxIVA algorithm based on this particular source prior is derived. The proposed method will be shown to better preserve the relationships between different frequency bins, thereby improving the separation performance.

The paper is organized as follows, in Section 2, the particular source prior is analyzed. In Section 3, the AuxIVA algorithm with our source prior is derived and the advantage of this source prior is discussed. Then, the experimental results are shown to confirm the advantages of the proposed method

in Section 4. Finally, conclusions are drawn in Section 5.

2. A MULTIVARIATE GENERALIZED GAUSSIAN DISTRIBUTION SOURCE PRIOR

For the convolutive blind source separation problem, the basic noise free model in the frequency domain is described as:

$$\mathbf{x}(k) = H(k)\mathbf{s}(k) \tag{1}$$

$$\mathbf{y}(k) = W(k)\mathbf{x}(k) \tag{2}$$

where $\mathbf{x}(k) = [x_1(k), \cdots, x_m(k)]^T$ is the observed signal vector, while $\mathbf{s}(k) = [s_1(k), \cdots, s_n(k)]^T$ and $\mathbf{y}(k) = [y_1(k), \cdots, y_n(k)]^T$ are the source signal vector and the estimated source vector respectively in the frequency domain; and $(\cdot)^T$ denotes vector transpose. $H(k)$ is the mixing matrix with $m \times n$ dimension, and $W(k)$ is the unmixing matrix with $n \times m$ dimension. The index k denotes the $k-th$ frequency bin, and $k = 1, \cdots, K$. The unmixing matrix can be defined as

$$W(k) = [\mathbf{w}_1(k) \cdots \mathbf{w}_n(k)]^h \tag{3}$$

where $(\cdot)^h$ denotes Hermitian transpose.

For traditional CBSS approaches, the scalar Laplace distribution is widely used for the source prior. However, the resultant nonlinear score function is a univariate function, which can not retain the dependency between different frequency bins for each source. In [8], IVA exploits a dependent multivariate Laplace distribution as the source prior. In this paper we adopt a particular multivariate distribution as the source prior which can be written as:

$$p(\mathbf{s}_i) \propto exp\Big(- \sqrt[3]{(\mathbf{s}_i - \mu_i)^h \Sigma_i^{-1} (\mathbf{s}_i - \mu_i)} \Big) \tag{4}$$

where $\mathbf{s}_i = [s_i(1), \cdots, s_i(K)]^T$ is the i-th vector source, and μ_i and Σ_i^{-1} are respectively the mean vector and covariance matrix. The covariance matrix is then assumed to be a diagonal matrix due to the orthogonality of the Fourier bases, which implies that each frequency bin sample is mutually uncorrelated. Moreover, if the source signal is zero mean and unity variance, then the source prior becomes:

$$p(\mathbf{s}_i) \propto exp\Big(- \Big(\sum_{k=1}^{K} |s_i(k)|^2 \Big)^{\frac{1}{3}} \Big) = exp\Big(- \Big(\|\mathbf{s}_i\|_2 \Big)^{\frac{2}{3}} \Big) \tag{5}$$

where $|\cdot|$ denotes the absolute value, and $\|\cdot\|_2$ denotes the Euclidean norm. This new source prior can be taken as a multivariate generalized Gaussian distribution with the shape parameter $\frac{2}{3}$, and the original multivariate Laplace distribution also belongs to this family with the shape parameter 1. For the multivariate generalized Gaussian distribution, the smaller the shape parameter, the heavier are the tails. Thus, this source prior has a heavier tail, which which can have advantage in separating non-stationary signals.

3. AUXIVA USING THE PARTICULAR SOURCE PRIOR

The auxiliary function technique is developed from the expectation-maximization algorithm, and avoids the step size tuning [14]. In the auxiliary function technique, an auxiliary function is designed for optimization. Instead of minimizing the cost function, the auxiliary function is minimized in terms of auxiliary variables. The auxiliary function technique can guarantee monotonic decrease of the cost function, and therefore provides effective iterative update rules [13].

The contrast function for AuxIVA is derived from the source prior [14]. For the original IVA algorithm,

$$G(\mathbf{y}_i) = G_R(r_i) = r_i \tag{6}$$

where $r_i = \|\mathbf{y}_i\|_2$.

By using the proposed source prior, we can obtain the following contrast function

$$G(\mathbf{y}_i) = G_R(r_i) = r_i^{\frac{2}{3}}. \tag{7}$$

The update rules contain two parts, i.e. the auxiliary variable updates and unmixing matrix updates. In summary, the update rules are as follow:

$$r_i = \sqrt{\sum_{k=1}^{K} |\mathbf{w}_i^h(k)\mathbf{x}(k)|^2} \tag{8}$$

$$V_i(k) = E[\frac{G'_R(r_i)}{r_i}\mathbf{x}(k)\mathbf{x}(k)^h] \tag{9}$$

$$\mathbf{w}_i(k) = (W(k)V_i(k))^{-1}\mathbf{e}_i \tag{10}$$

$$\mathbf{w}_i(k) = \frac{\mathbf{w}_i(k)}{\sqrt{\mathbf{w}_i^h(k)V_i(k)\mathbf{w}_i(k)}}. \tag{11}$$

In equation (10), $\mathbf{e}_i$ is a unity vector, the i-th element of which is unity.

During the update process of the auxiliary variable $V_i(k)$, we notice that $\frac{G'_R(r_i)}{r_i}$ is used to keep the dependency between different frequency bins for source i. In this paper, as we defined previously, $G_R(r_i) = r_i^{\frac{2}{3}}$. Therefore

$$\frac{G'_R(r_i)}{r_i} = \frac{2}{3r_i^{\frac{4}{3}}} = \frac{2}{3\sqrt[3]{(\sum_{k=1}^{K}|y_i(k)|^2)^2}} \tag{12}$$

If we expand the equation under the cubic root in equation (12), then

$$(\sum_{k=1}^{K}|y_i(k)|^2)^2 = \sum_{k=1}^{K}|y_i(k)|^4 + \sum_{u\neq v} c_{uv}|y_i(u)|^2|y_i(v)|^2 \tag{13}$$

which contains the cross items $\sum_{u\neq v} c_{uv}|y_i(u)|^2|y_i(v)|^2$, where c_{uv} denotes a scalar coefficient between the u-th and

v-th frequency bins. These terms contain the fourth order inter-relationships of different components for each source vector. Thus, they can provide an informative model of the dependency structure.

The use of such fourth order information has seldom been highlighted in original AuxIVA. We will show the comparison of the second order information and the fourth order information inherent to speech signals. We choose a particular speech signal "si10390.wav" from the TIMIT database [15], with 8 kHz sampling frequency and 1024 DFT length. Fig 1 is part of the covariance matrix, which is correspondent to the low frequency bins. It is difficult to observe any information in the high frequency bins due to the limited energy. We can see that the second order information is mainly distributed on the diagonal. This is because the Fourier transform is an orthogonal based transform.

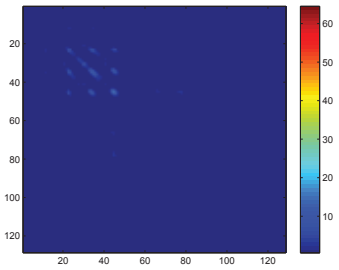


Fig. 1. Second order inter-frequency relationships information for the speech signal "si1039.wav", x and y dimensions correspond to frequency bins 1 to 128 of 512.

Now we construct a fourth order matrix to exploit the fourth order information

$$\begin{pmatrix} E[(y_i(1))^2(y_i(1))^2] & \cdots & E[(y_i(1))^2(y_i(K))^2] \\ \vdots & \ddots & \vdots \\ E[(y_i(K))^2(y_i(1))^2] & \cdots & E[(y_i(K))^2(y_i(K))^2] \end{pmatrix} \quad (14)$$

Fig 2 is part of this fourth order matrix, which is also correspondent to the same low frequency bins. It is evident that there is fourth order information not only on the diagonal. Thus, such fourth order information should be exploited to help separation, as have been recently highlighted in [16], namely considering the correlation of squares of components.

Now we will discuss that the proposed source prior is the most appropriate for introducing the fourth order relationship as shown in equation (13). We assume that the source prior has the general form

$$p(\mathbf{s}_i) \propto exp\Big(-\Big(\sum_{k=1}^{K}|s_i(k)|^2\Big)^{\beta}\Big) = exp\Big(-\Big(\|\mathbf{s}_i\|_2\Big)^{2\beta}\Big) \quad (15)$$

Thus the general contrast function is:

$$G(\mathbf{y}_i) = G_R(r_i) = r_i^{2\beta} \quad (16)$$

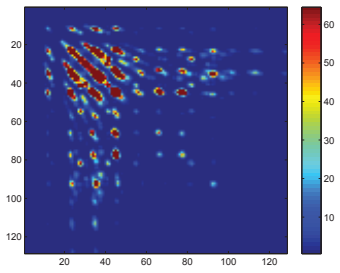


Fig. 2. Fourth order inter-frequency relationships information for the speech signal "si1039.wav", x and y dimensions correspond to frequency bins 1 to 128 of 512.

Then the general form of equation (12) is:

$$\frac{G'_R(r_i)}{r_i} = \frac{2\beta}{r_i^{2(1-\beta)}} = \frac{2\beta}{(\sum_{k=1}^{K}|y_i(k)|^2)^{1-\beta}} \quad (17)$$

In order to preserve the form of the fourth order relationship as defined in equation (13), the root needs to be odd. Thus the following condition must be satisfied

$$1-\beta = \frac{2}{2I+1} \quad (18)$$

where I is positive integer. Then we can obtain the condition for β

$$\beta = \frac{2I-1}{2I+1} \quad (19)$$

On the other hand, β is the shape parameter of the generalized multivariate Gaussian distribution. In order to make the proposed source prior more robust to outliers, β should be less than the 1/2, which is correspondent to the original source prior. Thus

$$\frac{2I-1}{2I+1} < \frac{1}{2} \quad (20)$$

Finally, $I = 1$ is the only solution, and the correspondent β is 1/3, as proposed in this paper.

4. EXPERIMENTS AND RESULTS

In this section, we use experiments to confirm the advantages of the proposed method. We chose different speech signals from the TIMIT dataset [15]. Each speech signal was approximately seven seconds long. The image method was used to generate the room impulse responses, and the size of the room was $7 \times 5 \times 3m^3$. The DFT length was 1024. We used a 2×2 mixing case, i.e. $m = n = 2$, for which the microphone positions are [3.48, 2.50, 1.50]m and [3.52, 2.50, 1.50]m respectively. The sampling frequency was 8kHz. The separation performance was evaluated objectively by the signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR)[17]. Fig 3 denotes the experimental setting.

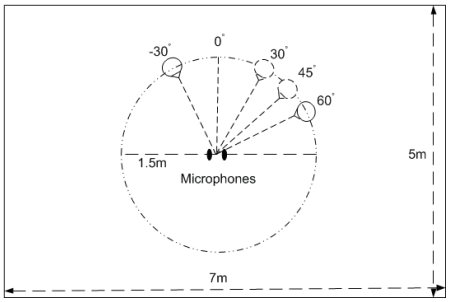


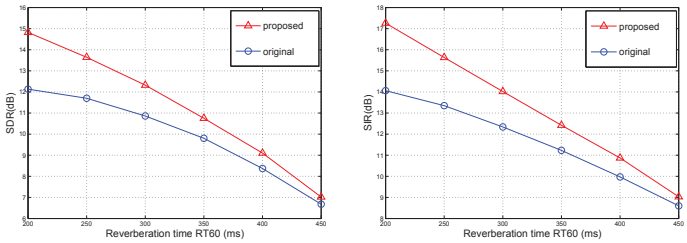
Fig. 3. Plan view of the experiment setting in the room environment with two microphones and two sources

In the first experiment, we set the reverberation time RT60 = 200ms. We selected two different speech signals randomly from the TIMIT dataset and convolved them into two mixtures. Then the orignal AuxIVA method and the proposed AuxIVA method with the new source prior were used to separate the mixtures respectively. Then we changed the source positions to repeat the simulation. For every pair of speech signals, three different azimuth angles for the sources relative to the normal to the microphone array were set for testing, these angles were selected from 30, 45, 60 and -30 degrees as shown in Fig 3. After that, we chose another pair of speech signals to repeat the above simulations. In total, we used five different pairs of speech signals, and repeated the simulation 15 times at different positions. Table 1 shows the average separation performance for each pair of speech signals in terms of SDR and SIR. The average SDR and SIR improvements are approximately 1.7dB and 1.9dB respectively. The results confirm the advantage of the proposed AuxIVA method which can better preserve the dependency between different frequency bins of each source.

Table 1. Separation performance comparison in terms of SDR and SIR measures in dB

Mixtures	mix1	mix2	mix3	mix4	mix 5
AuxIVA (SDR)	12.13	14.62	9.86	19.23	18.64
Proposed (SDR)	14.82	16.30	12.45	19.92	19.50
AuxIVA (SIR)	14.06	16.72	11.59	20.54	20.12
Proposed (SIR)	17.26	18.42	14.58	21.20	20.90

In the second experiment, we tested the robustness of the proposed AuxIVA method in different reverberant room environments. We selected two speech signals from the TIMIT dataset randomly and convolved them into two mixtures. The azimuth angles for the sources relative to the normal to the microphone array were set as 60 and -30 degrees. Both the original AuxIVA and the proposed AuxIVA were used to separate the mixtures. The results are shown in Fig 4, which shows the separation performance comparison in different reverberant environments. Fig 4(a) and Fig 4(b) show the SDR and SIR comparison respectively. It indicates that the proposed algorithm can consistently improve the separation performance in different reverberant environment.



(a) (b)

Fig. 4. Separation comparison between original and proposed AuxIVA algorithms as a function of reverberation time

5. CONCLUSIONS

The auxiliary function based IVA algorithm is a recently proposed fast form IVA algorithm. For all the IVA algorithms, the selection of the source prior is important. In this paper, we examined a particular multivariate generalized Gaussian distribution as the source prior, then the AuxIVA method was derived based on this particular source prior. The proposed method exploits a score function which contains a term describing the fourth order relationships between different frequency bins for each source, and thereby provides a more informative dependency structure. The experimental results showed that the proposed AuxIVA method can improve the separation performance by approximately 1.7dB and 1.9dB respectively in terms of SDR and SIR. Meanwhile, the proposed method can consistently improve the separation performance in different reverberant environments.

6. REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: learning algorithms and applications*, Wiley, 2003.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, 2001.

[3] C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of The Acoustical Society of America*, vol. 25, pp. 975–979, 1953.

[4] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Computation*, vol. 17, pp. 1875–1902, 2005.

[5] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 320–327, 2000.

[6] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook on Speech Processing and Speech Communication*, 2007.

[7] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: definition and algorithms," in *Fortieth Asilomar Conference on Signals, Systems and Computers 2006*, Asilomar, USA, 2006.

[8] T. Kim, H. Attias, S. Lee, and T. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech and Language processing*, vol. 15, pp. 70–79, 2007.

[9] T. Kim, "Real-time independent vector analysis for convolutive blind source separation," *IEEE Transactions on Circuits and Systems I*, vol. 57, pp. 1431–1438, 2010.

[10] Y. Liang, S.M. Naqvi, and J. Chambers, "Adaptive step size indepndent vector analysis for blind source separation," in *17th International Conference on Digital Signal Processing*, Corfu, Greece, 2011.

[11] I. Lee, T. Kim, and T.-W. Lee, "Fast fixed-point independent vector analysis algorithms for convolutive blind source separation," *Signal Processing*, vol. 87, pp. 1859–1871, 2007.

[12] Y. Liang, S.M. Naqvi, and J. Chambers, "Audio video based fast fixed-point independent vector analysis for multisource separation in a room environment," *EURASIP Journal on Advances in Signal Processing*, vol. 2012:183, 2012.

[13] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *2011 IEEE WASPAA*, New Paltz, USA, 2011.

[14] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian source," in *LVA/IVA 2010*, St. Malo, France, 2010.

[15] J. S. Garofolo et al., "TIMIT acoustic-phonetic continuous speech corpus," in *Linguistic Data Consortium*, Philadelphia, 1993.

[16] A. Hyvärinen, "Independent component analysis: recent advances," *Philos Transact A Math Phys Eng Sci*, vol. 371(1984), pp. 1–19, 2012.

[17] E. Vincent, C. Févotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 1462–1469, 2006.