Chapter 5
Frequency Domain Blind Source Separation Based on Independent Vector Analysis with a Multivariate Generalized Gaussian Source Prior

Yanfeng Liang, Syed Mohsen Naqvi, Wenwu Wang and Jonathon A. Chambers

Abstract Independent vector analysis (IVA) is designed for retaining the dependency contained in each source vector, while removing the dependency between different source vectors during the source separation process. It can theoretically avoid the permutation problem inherent to independent component analysis (ICA). The dependency in each source vector is maintained by adopting a multivariate source prior instead of a univariate source prior. In this chapter, a multivariate generalized Gaussian distribution is proposed to be the source prior, which can exploit the energy correlation within each source vector. It can preserve the dependency between different frequency bins better to achieve an improved separation performance, and is suitable for the whole family of IVA algorithms. Experimental results on real speech signals confirm the advantage of adopting the new source prior on three types of IVA algorithms.

5.1 Introduction

Blind source separation (BSS) aims to separate specific signals from observed mixtures with very limited prior knowledge, and has been researched over recent decades and has wide potential applications, such as in biomedical signal processing, image

Y. Liang (✉) · S. M. Naqvi · J. A. Chambers
Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK
e-mail: y.liang2@lboro.ac.uk

S. M. Naqvi
e-mail: S.M.R.Naqvi@lboro.ac.uk

J. A. Chambers
e-mail: J.A.Chambers@lboro.ac.uk

Wenwu Wang
University of Surrey, Guildford, Surrey, GU2 7XH, UK
e-mail: W.Wang@surrey.ac.uk

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_5,

processing, speech processing, and communication systems [1, 2]. A classical BSS problem is the machine cocktail party problem, which was proposed by Colin Cherry in 1953 [3, 4]. His drive was a machine to mimic the ability of a human to extract a target speech signal from microphone measurements acquired in a room environment.

In order to solve the BSS problem, a statistical signal processing method, i.e., independent component analysis (ICA), is proposed to exploit the non-Gaussianity of the signals [5]. It works efficiently to solve the instantaneous BSS problem. However, the problem becomes convolutive BSS problem in a room environment due to the reflections from the ceiling, floor, and walls. The length of the room impulse response is typically on the order of thousands of samples, which leads to huge computational cost when using time domain methods. Therefore, frequency domain methods have been proposed to reduce the computational cost due to the convolution operation in the time domain becomes multiplication in the frequency domain provided the block length of the transform is substantially larger than the length of the time domain filter [6, 7]. When the mixtures are transformed into the frequency domain by using the discrete Fourier transform (DFT), the instantaneous ICA can be applied in each frequency bin to separate the signals. However, the permutation ambiguity inherent to ICA becomes more severe because of the potential misalignment of the separated sources at different frequency bins. In this case, when the separated sources are transformed back to the time domain, the separation performance will be poor. Therefore, various methods have been proposed to mitigate the permutation problem [7]. However, most of them use extra information such as source geometry or prior knowledge of the source structure, and pre or post processing is needed for all of these methods which introduces additional complexity and delay.

Recently, independent vector analysis (IVA) has been proposed to solve the permutation problem naturally during the learning process without any pre or post processing [8]. It can theoretically avoid the permutation problem by retaining the dependency in each individual source vector while removing the dependency between the source vectors of different signals [9, 10]. The main difference between ICA algorithms and IVA algorithms is the nonlinear score function. For conventional ICA algorithms, the nonlinear score function is a univariate function which only uses the data in each frequency bin to update the unmixing matrix. However, the nonlinear score function for IVA is a multivariate function, which can use the data from all the frequency bins. Therefore, it can exploit the inter-frequency dependencies to mitigate the permutation problem.

There are three state-of-the-art types of IVA algorithms, which are the natural gradient IVA (NG-IVA), the fast fixed-point IVA (FastIVA) and the auxiliary function based IVA (AuxIVA). NG-IVA adopts the Kullback-Leibler divergence between the joint probability density function and the product of marginal probability density functions of the individual source vectors as the cost function, and the natural gradient method is used to minimize the cost function [9]. FastIVA is a fast form of IVA which uses Newton's method to update the unmixing matrix [11]. AuxIVA uses the auxiliary function technique to converge quickly without introducing tuning parameters and can guarantee the objective function decreases monotonically [12]. There are also several other IVA algorithms, which are based on these three IVA algorithms. The adaptive step size IVA algorithm, which is based on the NG-IVA algorithm, can automatically select the step size to achieve a faster convergence [13]. The audio-video based IVA method combines video information with FastIVA to obtain a faster and better separation performance in noisy and reverberant room environments [14]. And IVA methods which exploit the activity and dynamic structure of the sources to achieve improved separation performance have also been proposed [15, 16].

The nonlinear score function of IVA is used to preserve inter-frequency dependencies for individual sources [9]. Because the nonlinear score function is derived from the source prior, an appropriate source model is needed. For the original IVA algorithms, a multivariate Laplace distribution is adopted as the source prior. It is a spherically symmetric distribution, which implies the dependencies between different frequency bins are all the same. In order to describe a variable dependency structure, a chain-type overlapped source model has been proposed [17]. Similarly, a harmonic structure dependency model has been proposed [18]. A Gaussian mixture model can also be adopted as the source prior, whose advantage is that it enables the IVA algorithms to separate a wider class of signals [19, 20]. Most of these source models assume the covariance matrix of each source vector is an identity matrix due to the orthogonal Fourier basis. This implies that there is no second order correlation between different frequency bins. Although recently a multivariate Gaussian source prior has been proposed to introduce the second order correlation [21], it is only applicable when there are large second order correlations such as in functional magnetic resonance imaging (fMRI) studies. For the frequency domain IVA algorithms, higher order correlation information between different frequency bins is still missing and should be exploited.

In this chapter, a multivariate generalized Gaussian distribution is adopted as the source prior. It has heavier tails compared with the original multivariate Laplacian distribution, which makes the IVA algorithms derived from it more robust to outliers. It can also preserve the dependency across different frequency bins in a similar way as when the original multivariate Laplacian distribution is used to derive an IVA algorithm. Moreover, the nonlinear source function derived from this new source prior can additionally introduce the energy correlation within each source vector. Therefore, it contains more informative dependency structure and can thereby better preserve the dependencies between different frequency bins to achieve an improved separation performance.

The structure of this chapter is as follows. In Sect. 5.2, the original IVA is introduced. In Sect. 5.3, the energy correlation within a frequency domain speech signal is introduced. Then a multivariate generalized Gaussian distribution is proposed to be the source prior and analyzed in Sect. 5.4. Three types of IVA algorithms with the proposed source prior are discussed in Sect. 5.5. The experimental results are shown in Sect. 5.6, and finally the conclusions are drawn in Sect. 5.7.

5.2 Independent Vector Analysis

In this chapter, we mainly focus on the IVA algorithms used in the frequency domain. The noise-free model in the frequency domain is described as:

$$\mathbf{x}(k) = \mathbf{H}(k)\mathbf{s}(k) \tag{5.1}$$

$$\hat{\mathbf{s}}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{5.2}$$

where $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ is the observed signal; $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is the source signal; $\hat{\mathbf{s}}(k) = [\hat{s}_1(k), \ldots, \hat{s}_n(k)]^T$ is the estimated signal. They are all in the frequency domain and $(\cdot)^T$ denotes vector transpose. The index $k = 1, 2, \ldots, K$ denotes the k-th frequency bin, and K is the number of frequency bins; m is the number of microphones and n is the number of sources. $\mathbf{H}(k)$ is the mixing matrix whose dimension is $m \times n$, and $\mathbf{W}(k)$ is the unmixing matrix whose dimension is $n \times m$. In this chapter, we assume that the number of sources is the same as the number of microphones, i.e., $m = n$.

Independent vector analysis is proposed to avoid the permutation problem by retaining the inter-frequency dependencies for each source while removing the dependencies between different sources. It theoretically ensures that the alignment of the separated signals are consistent across the frequency bins. The IVA method adopts the Kullback-Leibler divergence [9] between the joint probability density function $p(\hat{\mathbf{s}}_1 \ldots \hat{\mathbf{s}}_n)$ and the product of marginal probability density functions of the individual source vectors $\prod q(\hat{\mathbf{s}}_i)$ as the cost function

$$\begin{aligned} J &= KL\Big(p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right) || \prod q\left(\hat{\mathbf{s}}_i\right)\Big) \\ &= \int p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right) \log \frac{p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right)}{\prod q\left(\hat{\mathbf{s}}_i\right)} d\hat{\mathbf{s}}_1 \cdots d\hat{\mathbf{s}}_n \\ &= \text{const} - \sum_{k=1}^{K} \log|\det\left(\mathbf{W}(k)\right)| - \sum_{i=1}^{n} E\left[\log q\left(\hat{\mathbf{s}}_i\right)\right] \end{aligned} \tag{5.3}$$

where $E[\cdot]$ denotes the statistical expectation operator, and $\det(\cdot)$ is the matrix determinant operator. The dependency between different source vectors should be removed but the inter-relationships between the components within each vector can be retained, when the cost function is minimized. These inter-frequency dependencies are modelled by the probability density function of the source.

Traditionally, the scalar Laplacian distribution is widely used as the source prior for the frequency domain ICA-based approaches. The resultant nonlinear score function is a univariate function, which can not preserve the inter-frequency dependencies because it is only associated with each individual frequency bin. In order to keep the inter-frequency dependencies of each source vector, a multivariate Laplacian distribution is adopted as the source prior for IVA, which can be written as

$$q(\mathbf{s}_i) \propto \exp\left(-\sqrt{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \tag{5.4}$$

where $(\cdot)^{\dagger}$ denotes the Hermitian transpose, $\boldsymbol{\mu}_i$ and Σ_i are respectively the mean vector and covariance matrix of the i-th source. Then the nonlinear score function can be derived according to this source prior. We assume that the mean vector is a zero vector and the covariance matrix is a diagonal matrix due to the orthogonality of the Fourier basis, which implies that each frequency bin sample is uncorrelated with the others. As such, the resultant nonlinear score function to extract the i-th source at the k-th frequency bin can be obtained as:

$$\begin{aligned}\varphi^{(k)}\left(\hat{s}_i(1)\cdots\hat{s}_i(k)\right) &= -\frac{\partial \log\left(q\left(\hat{s}_i(1)\cdots\hat{s}_i(k)\right)\right)}{\partial \hat{s}_i(k)} \\ = \frac{\partial\sqrt{\sum_{k=1}^{K}\left|\frac{\hat{s}_i(k)}{\sigma_i(k)}\right|^2}}{\partial \hat{s}_i(k)} &= \frac{\hat{s}_i(k)}{(\sigma_i(k))^2\sqrt{\sum_{k=1}^{K}\left|\frac{\hat{s}_i(k)}{\sigma_i(k)}\right|^2}}\end{aligned} \tag{5.5}$$

where $\sigma_i(k)$ denotes the standard deviation of the i-th source at the k-th frequency bin. This is a multivariate function, and the dependency between the frequency bins is thereby accounted for in learning. When the natural gradient method is used to minimize the cost function, the unmixing matrix update equation is:

$$\Delta\mathbf{W}(k) = \left(\mathbf{I} - E\left[\left(\varphi^{(k)}(\hat{\mathbf{s}})\right)\hat{\mathbf{s}}^{*}(k)\right]\right)\mathbf{W}(k) \tag{5.6}$$

where $\mathbf{I}$ is the identity matrix, and $(\cdot)^{*}$ denotes the conjugate operators. $\varphi^{(k)}(\hat{\boldsymbol{s}})$ is the nonlinear score function

$$\varphi^{(k)}(\hat{\boldsymbol{s}}) = \left[\varphi^{(k)}(\hat{\boldsymbol{s}}_{\mathbf{1}}), \ldots, \varphi^{(k)}(\hat{\boldsymbol{s}}_{\boldsymbol{n}})\right]^{T} \tag{5.7}$$

5.3 Energy Correlation Within a Frequency Domain Speech Signals

In the derivation of the original IVA algorithms little attention was focused upon the correlation information between different frequency bins due to the orthogonal Fourier basis. However, the higher order information exists and could be introduced to exploit the dependency between different frequency bins and better preserve the inter-frequency dependency. The correlation of squares of components is discussed in [22], which can be used to exploit the dependency between different components. For the frequency domain speech signals, the energy correlation between different

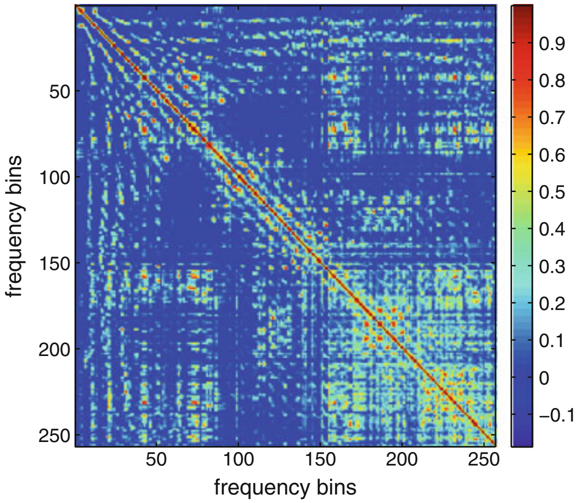

Fig. 5.1 The frequency domain energy correlation of the speech signal "si1039.wav"

frequency bins is such square correlation, which can be defined as

$$\mathrm{cov}\left(|s_i(a)|^2, |s_i(b)|^2\right) = E\left[|s_i(a)|^2|s_i(b)|^2\right] - E\left[|s_i(a)|^2\right]E\left[|s_i(b)|^2\right] \quad (5.8)$$

The use of such energy correlation has seldom been highlighted in the original IVA algorithms. In order to show the energy correlation within the frequency domain speech signals, we choose a particular speech signal "si10390.wav" from the TIMIT database [23], with 8 kHz sampling frequency and 1,024 DFT length. Then the matrix of energy correlation coefficients between different frequency bins is plotted as shown in Fig. 5.1. Figure 5.1 is just part of the whole matrix of energy coefficient, which corresponds to the frequency bins from 1 to 256. The high frequency part is omitted due to the limited energy which leads to large correlation coefficients.

It is shown in Fig. 5.1, besides the information on the diagonal, there are many information on the off-diagonal elements, which is correspond to the energy correlation between different frequency bins. It indicates that there are energy correlation as defines in Eq. (5.8), which also leads to that $E\left[|s_i(a)|^2|s_i(b)|^2\right]$ is not equal to zero for many points and this information should be used to help to further exploit the dependency between different frequency bins.

5.4 Multivariate Generalized Gaussian Source Prior

In this section, we propose a particular multivariate generalized Gaussian source prior as the source prior, from which a new nonlinear score function can be derived to introduce the energy correlation information to improve the separation performance.

The source prior we proposed belongs to the family of multivariate generalized Gaussian distributions which takes the form

$$q(\mathbf{s}_i) \propto \exp\left(-\left(\frac{(\mathbf{s}_i-\boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1}(\mathbf{s}_i-\boldsymbol{\mu}_i)}{\alpha}\right)^{\beta}\right) \tag{5.9}$$

when $\alpha = 1$ and $\beta = \frac{1}{2}$, it is the multivariate Laplace distribution adopted by the original IVA algorithm [9].

Now we assume that $\alpha = 1$, the mean vector is a zero vector and the covariance matrix is an identity matrix due to the orthogonality of the Fourier basis and scaling adjustment. Then, the source prior takes the general form

$$p(\mathbf{s}_i) \propto \exp\left(-\left(\sum_{k=1}^{K}|s_i(k)|^2\right)^{\beta}\right) \tag{5.10}$$

where we constrain $0 < \beta < 1$ to obtain a super-Gaussian distribution to describe the speech signals. The nonlinear score function based on this new source prior is

$$\varphi^{(k)}\left(\hat{s}_i(1)\ldots\hat{s}_i(k)\right) = \frac{2\beta\hat{s}_i(k)}{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^{1-\beta}} = \frac{2\beta\hat{s}_i(k)}{\left(\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2\right)^{\frac{1-\beta}{2}}} \tag{5.11}$$

In order to introduce the energy correlation, the root needs to be odd, otherwise the square will be cancelled. Then the denominator can be expanded as

$$\left(\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2\right)^{\frac{1-\beta}{2}} = \left(\sum_{k=1}^{K}|\hat{s}_i(k)|^4 + \sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2\right)^{\frac{1-\beta}{2}} \tag{5.12}$$

which contains cross items $\sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2$ corresponding to energy correlation between different frequency bins, and c_{ab} is a scalar constant between the a-th and b-th frequency bins.

Thus the following condition must be satisfied

$$\frac{1-\beta}{2} = \frac{1}{2I+1} \tag{5.13}$$

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

$$\beta = \frac{2I-1}{2I+1} \tag{5.14}$$

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 compared with the original source prior, β should be less than the 1/2, which is correspondent to the original source prior. Thus

$$\frac{2I-1}{2I+1} < \frac{1}{2} \tag{5.15}$$

Finally, $I = 1$ is the only solution, and the associated β is 1/3. Thus the appropriate generalized Gaussian distribution takes the form

$$q(\mathbf{s}_i) \propto \exp\left(-\sqrt[3]{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \tag{5.16}$$

We next show that this source prior can also preserve the inter-frequency dependencies within each source vector in a similar manner to the original source prior for IVA [9].

We begin with the definition of a K-dimensional random variable

$$\mathbf{s}_i = v^{\frac{3}{4}}\boldsymbol{\xi}_i + \boldsymbol{\mu}_i \tag{5.17}$$

where v is a scalar random variable, and $\boldsymbol{\xi}_i$ obeys a generalized Gaussian distribution which has the form:

$$p(\boldsymbol{\xi}_i) \propto \exp\left(-\left(\frac{\boldsymbol{\xi}_i^{\dagger}\Sigma_i^{-1}\boldsymbol{\xi}_i}{2\sqrt{2}}\right)^{\frac{2}{3}}\right). \tag{5.18}$$

If v has a Gamma distribution of the form:

$$p(v) \propto v^{\frac{1}{2}} \exp\left(-\frac{v}{2}\right) \tag{5.19}$$

then the proposed source prior can be achieved by integrating the joint distribution of $\mathbf{s}_i$ and v over v as follows:

$$\begin{aligned} q(\mathbf{s}_i) &= \int_0^{\infty} q(\mathbf{s}_i|v)p(v)dv \\ &= \alpha_1 \int_0^{\infty} v^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{\left((\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)\right)^{\frac{2}{3}}}{v} + v\right)\right) dv \\ &= \alpha_2 \exp\left(-\sqrt[3]{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \end{aligned} \tag{5.20}$$

where α_1 and α_2 are both normalization terms. Therefore, equation (5.20) confirms that the new source prior has the dependency generated by v.

In Lee's paper [24], the source priors suitable for IVA are discussed. A general form of source prior is described as:

$$q(\mathbf{s}_i) \propto \exp\left(-(\|\mathbf{s}_i\|_p)^{\frac{1}{L}}\right) = \exp\left(-\left(\sum_k |s_i^{(k)}|^p\right)^{\frac{1}{pL}}\right) \tag{5.21}$$

where $\|\cdot\|_p$ denotes the l_p norm, and L is termed as the sparseness parameter. He suggested that the spherical symmetry assumption is suitable for modeling the frequency components of speech, i.e. $p = 2$, and through certain experimental studies found that the best separation performance can be achieved when $L = 7$.

Our new proposed source prior also belongs to this family. If we choose $p = 2$ to make it spherically symmetric, and choose $L = \frac{3}{2}$, the proposed source prior can be obtained. However, our detailed experimental results show that the improvement of performance is not robust when $L = 7$ as mentioned in [24], while the NG-IVA which adopts our new source prior can consistently achieve improved separation performance.

5.5 IVA Algorithms with the New Source Prior

5.5.1 NG-IVA with the New Source Prior

Applying this new source prior to derive the nonlinear score function of the NG-IVA algorithm with the assumption that the mean vector is zero and the covariance matrix is an identity matrix, we can obtain

$$\varphi^{(k)}\left(\hat{s}_i(1)\ldots\hat{s}_i(k)\right) = \frac{2\hat{s}_i(k)}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}}. \tag{5.22}$$

If we expand the equation under the cubic root, it can be written as:

$$\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2 = \sum_{k=1}^{K}|\hat{s}_i(k)|^4 + \sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2 \tag{5.23}$$

which contains cross items $\sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2$. These terms are related to the energy correlation between different components within each source vector, and capture the level of interdependency between different frequency bins. Thus, this new multivariate nonlinear function can provide a more informative model of the dependency structure. Moreover, it can better describe the speech model.

5.5.2 FastIVA with the New Source Prior

Fast fixed-point independent vector analysis is a fast form of IVA algorithm. Newton's method is adopted in the update, which converges quadratically and is free from selecting an efficient learning rate [11]. The contrast function used by FastIVA is as follows:

$$J = \sum_{i=1}^{n} \left(E\left[F\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \right] - \sum_{k=1}^{K} \lambda_i^{(k)} \left(\mathbf{w}_i(k)^\dagger \mathbf{w}_i(k) - 1 \right) \right) \tag{5.24}$$

where $\mathbf{w}_i^\dagger$ is the i-th row of the unmixing matrix $\mathbf{W}$, and λ_i is the i-th Lagrange multiplier. $F(\cdot)$ is the nonlinear function, which can take on several different forms as discussed in [11]. It is a multivariate function of the summation of the desired signals in all frequency bins. With normalization, the learning rule is:

$$\begin{aligned} \mathbf{w}_i(k) \leftarrow & E\left[F'\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) + |\hat{s}_i(k)|^2 F''\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \right] \mathbf{w}_i(k) \\ & - E\left[(\hat{s}_i(k))^* F'\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \mathbf{x}^k \right] \end{aligned} \tag{5.25}$$

where $F'(\cdot)$ and $F''(\cdot)$ denote the derivative and second derivative of $F(\cdot)$, respectively. If this is used for all sources, an unmixing matrix $\mathbf{W}(k)$ can be constructed, which must be decorrelated with

$$\mathbf{W}(k) \leftarrow \left(\mathbf{W}(k)(\mathbf{W}(k))^\dagger \right)^{-1/2} \mathbf{W}(k). \tag{5.26}$$

When the multivariate Laplacian distribution is used as the source prior for the FastIVA algorithm, with the zero mean and unity variance assumptions, the nonlinear function takes the form

$$F\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) = \left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right)^{\frac{1}{2}}. \tag{5.27}$$

When the new multivariate generalized Gaussian distribution is used as the source prior, with the same assumptions, the nonlinear function becomes:

$$F\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)=\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^{\frac{1}{3}}. \tag{5.28}$$

Therefore, the first derivative becomes:

$$F'\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)=\frac{2}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}}. \tag{5.29}$$

It is very similar to Eq. (5.22), and it also contains cross terms which can exploit the energy correlation between different frequency bins. Thus, the FastIVA algorithm with the new source prior is likely to help improve the separation performance.

5.5.3 AuxIVA with the New Source Prior

AuxIVA adopts the auxiliary function technique to avoid the step size tuning [25]. In the auxiliary function technique, an auxiliary function is designed for optimization. During the learning process, 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 [12].

The contrast function for AuxIVA is derived from the source prior [25]. For the original IVA algorithm,

$$G(\hat{\mathbf{s}}_i)=G_R(r_i)=r_i \tag{5.30}$$

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

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

$$G(\hat{\mathbf{s}}_i)=G_R(r_i)=r_i^{\frac{2}{3}}. \tag{5.31}$$

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

$$r_i=\sqrt{\sum_{k=1}^{K}|\mathbf{w}_i^{\dagger}(k)\mathbf{x}(k)|^2} \tag{5.32}$$

$$V_i(k)=E\left[\frac{G'{}_R(r_i)}{r_i}\mathbf{x}(k)\mathbf{x}(k)^{\dagger}\right] \tag{5.33}$$

$$\mathbf{w}_i(k) = \Big(W(k)V_i(k)\Big)^{-1}\mathbf{e}_i \tag{5.34}$$

$$\mathbf{w}_i(k) = \frac{\mathbf{w}_i(k)}{\sqrt{\mathbf{w}_i^{\dagger}(k)V_i(k)\mathbf{w}_i(k)}}. \tag{5.35}$$

In Eq. (5.34), $\mathbf{e}_i$ 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 chapter, as we defined previously, $G_R(r_i) = r_i^{\frac{2}{3}}$. Therefore

$$\frac{G'_R(r_i)}{r_i} = \frac{2}{3r_i^{\frac{4}{3}}} = \frac{2}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}} \tag{5.36}$$

which has the same form as Eq. (5.29). The update rules also contain the fourth order terms to exploit the energy correlation within the frequency domain speech signal vectors and should thereby help to achieve a better separation performance.

5.6 Experiments

In this section, we used all three state-of-the-art IVA algorithms with the proposed multivariate generalized Gaussian source prior to separate the mixtures obtained in a reverberant room environment. The speech signals were chosen from the TIMIT dataset [23], and each of them was approximately 7 s long. The image method was used to generate the room impulse responses, and the dimensions of the room were 7 × 5 × 3 m^3. The DFT length was set to be 1,024, and the reverberation time RT60 $=$ 200 ms. We used a 2 × 2 mixing case, and the microphone positions are [3.48, 2.50, 1.50] and [3.52, 2.50, 1.50] m respectively. The sampling frequency was 8 kHz. The separation performance was evaluated objectively by the signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR) [26]. Figure 5.2 is the plan view of the experimental setting.

5.6.1 NG-IVA Algorithms Comparison

In the first experiment, two different speech signals were chosen randomly from the TIMIT dataset and were convolved into two mixtures. Then the NG-IVA method with original source prior, the NG-IVA method with our proposed source prior and NG-

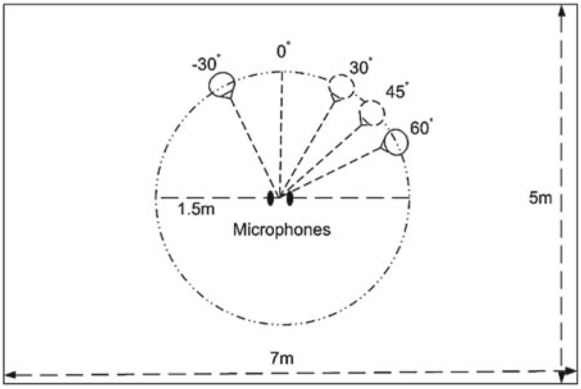

Fig. 5.2 Plan view of the experiment setting in the room environment with two microphones and two sources

Table 5.1 Separation performance comparison in SDR (dB)

	Original	Proposed	Lee's
Mixture 1	12.27	12.90	4.74
Mixture 2	18.13	18.47	18.34
Mixture 3	8.88	11.83	11.41
Mixture 4	15.57	16.92	5.95
Mixture 5	18.10	18.69	15.44
Mixture 6	18.81	19.58	3.71
Mixture 7	15.94	16.59	8.63
Mixture 8	15.29	15.75	16.03
Mixture 9	18.58	19.05	17.35
Mixture 10	18.80	19.31	0.78

IVA with Lee's source prior where the sparseness parameter $L = 7$, were all used to separate the mixtures, respectively. Then the source positions were changed 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 $-30°$. After that, we chose another pair of speech signals to repeat the above simulations. We used ten different pairs of speech signals totally, and repeated the simulation 30 times at different positions. Tables 5.1 and 5.2 show the average separation performance for each pair of speech signals in terms of SDR and SIR in dB.

The experimental results indicate clearly that NG-IVA with the proposed source prior can consistently improve the separation performance. Although the NG-IVA with Lee's source prior can get improvement results sometimes, the separation improvement is not consistent, in some cases there is essentially no separation such as mixtures 1, 6, and 10. Even though it can achieve better separation than original NG-IVA, it is still no better than the proposed method. Only for mixture 8, does

Table 5.2 Separation performance comparison in SIR (dB)

	Original	Proposed	Lee's
Mixture 1	14.08	14.84	5.62
Mixture 2	19.57	19.86	19.81
Mixture 3	10.72	13.74	13.19
Mixture 4	16.98	18.46	7.16
Mixture 5	20.14	20.47	16.94
Mixture 6	20.30	20.98	4.35
Mixture 7	17.88	18.40	10.73
Mixture 8	19.88	20.41	20.61
Mixture 9	20.75	20.89	18.80
Mixture 10	20.28	20.60	1.48

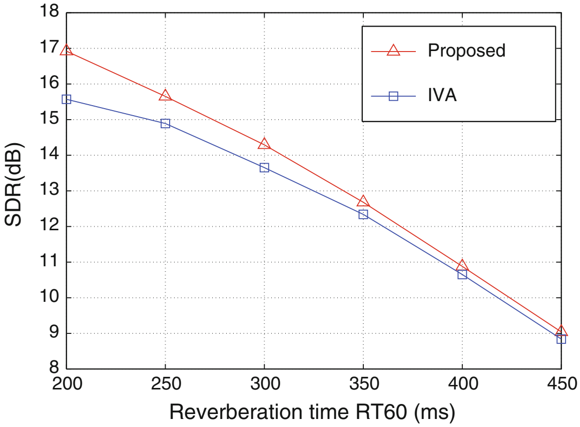

Fig. 5.3 Separation comparison in terms of SDR between original and proposed NG-IVA algorithms as a function of reverberation time

it achieve the best separation performance. Therefore, among all these three algorithms, the NG-IVA with the proposed source prior is the best method, because it can consistently achieve better separation performance. The average SDR improvement and SIR improvement are approximately 0.9 and 0.8 dB, respectively compared with the original NG-IVA algorithm.

Then we used the NG-IVA algorithms with the proposed source prior to separate the mixtures obtained from different reverberant room environments. Two speech signals were selected from the TIMIT dataset randomly and convolved into two mixtures. The azimuth angles for the sources relative to the normal to the microphone array were set as 60° and −30°. Both the original NG-IVA and the proposed method were used to separate the mixtures. The results are shown in Figs. 5.3 and 5.4, which show the separation performance comparisons in different reverberant environments. Figures 5.3 and 5.4 show the SDR and SIR comparison, respectively.

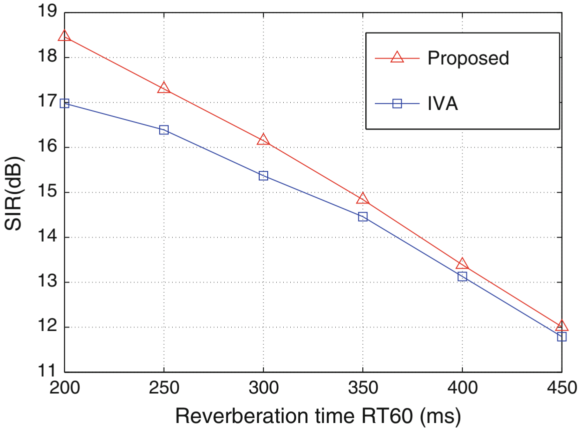

Fig. 5.4 Separation comparison in terms of SIR between original and proposed NG-IVA algorithms as a function of reverberation time

They indicate that the proposed algorithm can consistently improve the separation performance in different reverberant environments, up to a reverberation time of 450 ms. The advantage reduces with increasing reverberation time RT60 due to the greater challenge in extracting the individual source vectors.

5.6.2 FastIVA Algorithms Comparison

In the second experiment, all the experimental settings and the processes are all the same as the first experiment. Here we randomly selected five pairs of speech signals from the TIMIT dataset and convolved them into mixtures. The original FastIVA algorithm and the FastIVA algorithm with the proposed source prior were used to separate the speech mixtures. Then the source positions were changed to repeat the experiment, the average separation performance comparison is shown in Table 5.3. It shows that the separation performance can be improved by adopting the proposed source prior. The average SDR improvement and SIR improvement both are approximately 0.6 dB.

We also compared the separation performance of these two algorithms in different reverberant room environments as in the first experiment. The SDR and SIR comparisons are shown in Figs. 5.5 and 5.6. in terms of SDR and SIR comparison, respectively. The results show that the FastIVA algorithm with the proposed source prior can improve the separation performance, but again the advantage is reduced with increasing reverberation time RT60.

Table 5.3 Separation performance comparison in terms of SDR and SIR measures in dB

Mixtures	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5
Original FastIVA (SDR)	17.77	19.48	14.75	18.12	16.79
Proposed FastIVA (SDR)	18.04	20.63	15.08	18.88	17.31
Original FastIVA (SIR)	19.32	21.01	17.04	19.80	19.18
Proposed FastIVA (SIR)	19.59	22.04	17.31	20.51	19.74

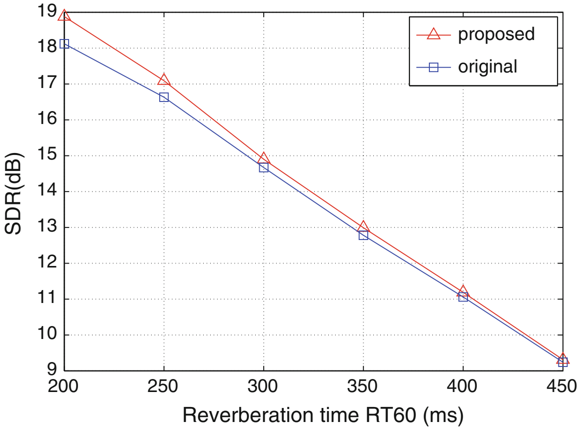

Fig. 5.5 Separation comparison in terms of SDR between original and proposed FastIVA algorithms as a function of reverberation time

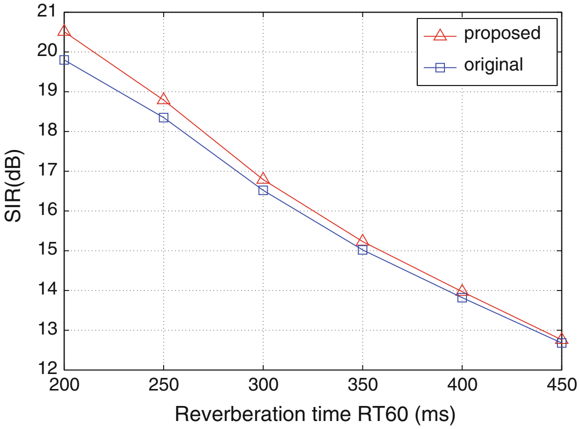

Fig. 5.6 Separation comparison in terms of SIR between original and proposed FastIVA algorithms as a function of reverberation time

Table 5.4 Separation performance comparison in terms of SDR and SIR measures in dB

Mixtures	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5
Original AuxIVA (SDR)	12.13	14.62	9.86	19.23	18.64
Proposed AuxIVA (SDR)	14.82	16.30	12.45	19.92	19.50
Original AuxIVA (SIR)	14.06	16.72	11.59	20.54	20.12
Proposed AuxIVA (SIR)	17.26	18.42	14.58	21.20	20.90

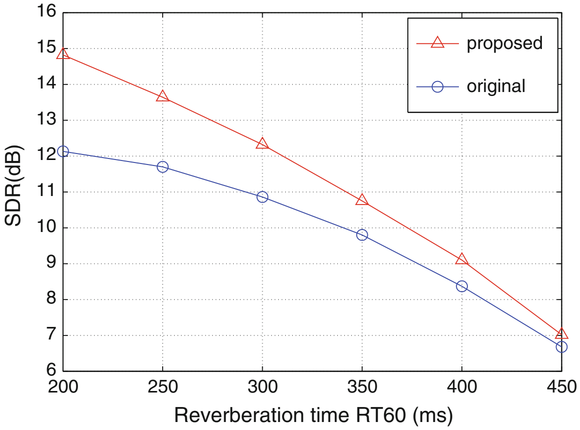

Fig. 5.7 Separation comparison in terms of SDR between original and proposed AuxIVA algorithms as a function of reverberation time

5.6.3 AuxIVA Algorithms Comparison

In the third experiment, the separation performance of AuxIVA with original source prior and AuxIVA with proposed source prior were compared. Again five different pairs of speech signals were used, and the simulation was repeated 15 times at different positions. Table 5.4 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 1.7 and 1.9 dB, respectively. The results confirm the advantage of the proposed AuxIVA method which can better preserve the dependency between different frequency bins of each source and thereby achieve a better separation performance.

Then we also tested the robustness of the proposed AuxIVA method in different reverberant room environments. The experimental settings are all the same as previous two experiments. The results are shown in Figs. 5.7 and 5.8, which show the separation performance comparison in different reverberant environments. It indicates that the AuxIVA algorithm with the proposed source prior can consistently improve the separation performance in different reverberant environments as the other two kinds of IVA algorithms.

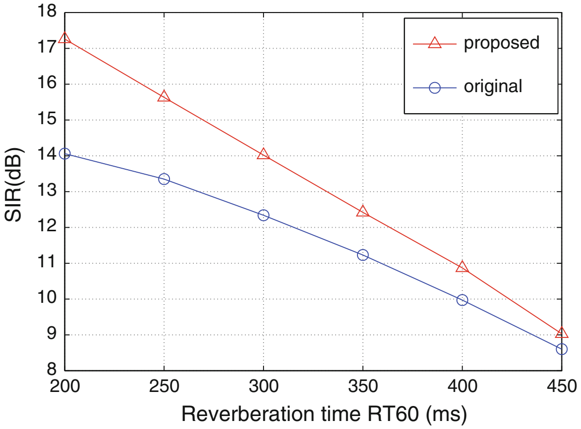

Fig. 5.8 Separation comparison in terms of SIR between original and proposed AuxIVA algorithms as a function of reverberation time

Examining the results for all the three algorithms, our proposed source prior offers the maximum improvement in the AuxIVA algorithm. However, it is difficult to make a general recommendation, which is the best algorithm due to the variability of performance with different speech signals and mixing environments.

5.7 Conclusions

In this chapter, a specific multivariate generalized Gaussian distribution was adopted as the source prior for IVA. This new source prior can better preserve the inter-frequency dependencies as compared to the original multivariate Laplace source prior, and is more robust to outliers. When the proposed source prior was used in IVA algorithms, it introduces energy correlation commonly found in frequency domain speech signals to improve the learning process and enhance separation. Three state-of-the-art types of IVA algorithms with the new source prior, i.e., NG-IVA, FastIVA, and AuxIVA, were all analyzed, and the experimental results confirm the advantage of adopting the new source prior particularly for low reverberation environment.

Acknowledgments Some of the material of this chapter is under review for publication in *Signal Processing* as "Independent Vector Analysis with a Generalized Multivariate Gaussian Source Prior for Frequency Domain Blind Source Separation," in October 2013.

References

1. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. Wiley, New York (2003)
2. Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, Oxford (2009)
3. Cherry, C.: Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. **25**, 975–979 (1953)
4. Cherry, C., Taylor, W.: Some further experiments upon the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. **26**, 554–559 (1954)
5. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
6. Parra, L., Spence, C.: Convolutive blind separation of non-stationary sources. IEEE Trans. Speech Audio Process. **8**, 320–327 (2000)
7. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. In: Handbook on Speech Processing and Speech Communication. Springer, New York (2007)
8. Kim, T., Lee, I., Lee, T.-W.: Independent vector analysis: definition and algorithms. In: Fortieth Asilomar Conference on Signals, Systems and Computers 2006. Asilomar, USA (2006)
9. Kim, T., Attias, H., Lee, S., Lee, T.: Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Audio Speech Lang. Process. **15**, 70–79 (2007)
10. Kim, T.: Real-time independent vector analysis for convolutive blind source separation. IEEE Trans. Circuits Syst. **57**, 1431–1438 (2010)
11. Lee, I., Kim, T., Lee, T.-W.: Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. Signal Process. **87**, 1859–1871 (2007)
12. Ono, N.: Stable and fast update rules for independent vector analysis based on auxiliary function technique. In: 2011 IEEE WASPAA. New Paltz, USA (2011)
13. Liang, Y., Naqvi, S.M., Chambers, J.: Adaptive step size indepndent vector analysis for blind source separation. In: 17th International Conference on Digital Signal Processing. Corfu, Greece (2011)
14. Liang, Y., Naqvi, S.M., Chambers, J.: Audio video based fast fixed-point independent vector analysis for multisource separation in a room environment. EURASIP J. Adv. Signal Process. **2012**, 183 (2012)
15. Masnadi-Shirazi, A., Zhang, W., Rao, B.D.: Glimpsing IVA: A framework for overcomplete/complete/undercomplete convolutive source separation. IEEE Trans. Audio Speech Lang. Process. **18**, 1841–1855 (2010)
16. Ono, T., Ono, N., Sagayama, S.: User-guided indpendent vector analysis with source activity tuning. In: ICASSP 2012. Kyoto, Japan (2012)
17. Lee, I., Jang, G.-J., Lee, T.-W.: Independent vector analysis using densities represented by chain-like overlapped cliques in graphical models for separation of convolutedly mixed signals. Electron. Lett. **45**, 710–711 (2009)
18. Choi, C.H., Chang, W., Lee, S.Y.: Blind source separation of speech and music signals using harmonic frequency dependent independent vector analysis. Electron. Lett. **48**, 124–125 (2012)
19. Lee, I., Hao, J., Lee, T.W.: Adaptive independent vector analysis for the separation of convoluted mixtures using EM algorithm. In: ICASSP 2008. USA, Las Vegas (2008)
20. Hao, J., Lee, I., Lee, T.W.: Independent vector analysis for source separation using a mixture of Gaussian prior. Neural Comput. **22**, 1646–1673 (2010)
21. Anderson, M., Adali, T., Li, X.-L.: Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis. IEEE Trans. Signal Process. **60**, 1672–1682 (2012)
22. Hyvärinen, A.: Independent component analysis: recent advances. Philos. Transact. A Math. Phys. Eng. Sci. **371**(1984), 1–19 (2012)
23. Garofolo, J.S., et al.: TIMIT acoustic-phonetic continuous speech corpus. Linguistic Data Consortium, Philadelphia (1993)

24. Lee, I., Lee, T.W.: On the assumption of spherical symmetry and sparseness for the frequency-domain speech model. IEEE Trans. Audio Speech Lang. Process. **15**, 1521–1528 (2007)
25. Ono, N., Miyabe, S.: Auxiliary-function-based independent component analysis for super-Gaussian source. In: LVA/IVA 2010. St. Malo, France (2010)
26. Vincent, E., Fevotte, C., Gribonval, R.: Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**, 1462–1469 (2006)