EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Independent vector analysis based on overlapped cliques of variable width for frequency-domain blind signal separation

Intae Lee[1] and Gil-Jin Jang[2*]

Abstract

A novel method is proposed to improve the performance of independent vector analysis (IVA) for blind signal separation of acoustic mixtures. IVA is a frequency-domain approach that successfully resolves the well-known permutation problem by applying a spherical dependency model to all pairs of frequency bins. The dependency model of IVA is equivalent to a single clique in an undirected graph; a clique in graph theory is defined as a subset of vertices in which any pair of vertices is connected by an undirected edge. Therefore, IVA imposes the same amount of statistical dependency on every pair of frequency bins, which may not match the characteristics of real-world signals. The proposed method allows variable amounts of statistical dependencies according to the correlation coefficients observed in real acoustic signals and, hence, enables more accurate modeling of statistical dependencies. A number of cliques constitutes the new dependency graph so that neighboring frequency bins are assigned to the same clique, while distant bins are assigned to different cliques. The permutation ambiguity is resolved by overlapped frequency bins between neighboring cliques. For speech signals, we observed especially strong correlations across neighboring frequency bins and a decrease in these correlations with an increase in the distance between bins. The clique sizes are either fixed, or determined by the reciprocal of the mel-frequency scale to impose a wider dependency on low-frequency components. Experimental results showed improved performances over conventional IVA. The signal-to-interference ratio improved from 15.5 to 18.8 dB on average for seven different source locations. When we varied the clique sizes according to the observed correlations, the stability of the proposed method increased with a large number of cliques.

Keywords: blind signal separation (BSS), independent component analysis (ICA), independent vector analysis (IVA)

1 Introduction

When an audio signal is recorded by a microphone in a closed room, it reaches the microphone via not only a direct path, but also infinitely many reverberant paths. The source sound wave is delayed in time and its energy is absorbed by walls when it is delivered by a reverberant path. In order to make the problem practically tractable, the time delay is usually limited to a certain number by which the signal energy may almost disappear through repeated reflections. The signal recorded by a digital microphone can then be modeled by a discrete convolution of a finite impulse response (FIR) filter and the source signal [1-3]. When there are multiple microphones and multiple sources, each microphone recording is expressed by the sum of the convolutions of corresponding transfer functions and source signals [4-6] such that

$$\begin{aligned} x_j(t) &= \sum_{i=1}^{M} \sum_{\tau=0}^{T} a_{ji}(\tau) s_i(t-\tau) \\ &= \sum_{i=0}^{M} a_{ji}(t) * s_i(t), \quad j = 1, \ldots, N, \end{aligned} \tag{1}$$

where the integer numbers j, M, N, and T are, the microphone number, number of sources, number of microphones, and order of the FIR filter, respectively. The time-domain sequences $x_j(t)$ and $s_i(t)$ are the signals recorded by microphone j and generated by source i,

* Correspondence: gjang@unist.ac.kr
[2]Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea
Full list of author information is available at the end of the article

respectively, and $a_{ji}(t)$ is the coefficient at time t of the FIR filter for the transfer function from source i to microphone j; it is affected by the recording environment, including the source and microphone locations. To ensure that the linear transformation is invertible, the number of sources should be equal to the number of microphones, i.e., $N = M$ [4].

This type of problem is often called blind signal separation (BSS) because there is no assumption of the source characteristics. Many studies have been carried out to tackle BSS problems based on independent component analysis (ICA), which minimizes the statistical dependency among the output signals [4-8]. However, direct inversion of the time-domain mixing filter in Equation 1 is difficult and often leads to unstable solutions. To obtain a more stable convergence, the short-time Fourier transform (STFT) is used to convert the convolution in Equation 1 to multiplications in the frequency domain [5]:

$$X_j(\omega, k) = A_{ji}(\omega) S_i(\omega, k), \quad j = 1, \dots, N, \tag{2}$$

where ω is the center frequency of each component of STFT, and the complex values $X_j(\omega, k)$, $A_{ji}(\omega)$, and $S_i(\omega, k)$ are STFT components of $x_j(t)$, $a_{ji}(t)$, and $s_i(t)$, respectively. Note that another discrete time domain exists which is denoted by the dummy variable k. This is different from the real-time variable t, as each value of k corresponds to a frame of the STFT. The value of $A_{ji}(\omega)$ is assumed to be constant over time, so it is not a function of k. Because we use discrete STFT, the center frequency of each discretized frequency bin is expressed as $\omega_b = \frac{b}{B}\omega_{\max}$, where B is the total number of frequency bins, b denotes the frequency bin number, and $\omega_{\max}$ is the maximum frequency equivalent to half of the Nyquist sampling rate. This means that the frequency-domain BSS methods only consider the STFT components at the frequencies in [0 π] [5]. The components at frequencies in [-π 0] can be reconstructed perfectly because a real-valued time-domain signal has a conjugate symmetric Fourier series: $X(-\omega) = \bar{X}(\omega)$ for $\omega \in$ [0 π], where $\overline{X}(\omega)$ is the complex conjugate of $X(\omega)$. For a more compact notation, we rewrite Equation 2 as

$$\mathbf{x}^b[k] = \mathbf{A}^b \mathbf{s}^b[k] \quad b = 1, 2, \dots, B, \tag{3}$$

where $\mathbf{x}^b[k] = [X_1(\omega_b, k) \dots X_N(\omega_b, k)]^T$, $\mathbf{s}^b[k] = [S_1(\omega_b, k) \dots S_M(\omega_b, k)]^T$, and $\mathbf{A}^b$ is an $N \times M$ matrix whose (j, i)th element is $A_{ji}(\omega_b, k)$. Dealing with the signals in the frequency domain improves the performance since longer filter lengths are better handled in the frequency domain and the convolved mixture problem reduces to an instantaneous mixture problem in each frequency bin; this is expressed as

$$\mathbf{y}^b[k] = \mathbf{W}^b \mathbf{x}^b[k], \quad b = 1, 2, \dots, B, \tag{4}$$

where $\mathbf{y}^b[k]$ is a vector of M estimated independent sources and $\mathbf{W}^b$ is an $M \times N$ matrix. Ideally, when $\mathbf{W}^b = (\mathbf{A}^b)^{-1}$, we can perfectly reconstruct the original sources by $\mathbf{y}^b[k] = (\mathbf{A}^b)^{-1}\mathbf{A}^b\mathbf{s}^b[k] = \mathbf{s}^b[k]$. However, all frequency-domain ICA algorithms inherently suffer from permutation and scaling ambiguity because they assume different frequency components to be independent [4,9]. The instantaneous ICA may assign individual frequency bins of a single source to different outputs, so grouping the frequency components of individual source signals is required for the success of the frequency-domain BSS [10]. One of the simplest solutions is smoothing the frequency-domain filter [10-12] at the expense of performance because of the lost frequency resolution. There are other methods for colored signals, such as explicitly matching components with larger inter-frequency correlations of signal envelopes [13-15].

Recently, a method called independent vector analysis (IVA) has been developed to overcome the permutation problem by embedding statistical dependency across different frequency components [16-19]. The joint dependency model assumes that the frequency bins of the acoustic sources have radially symmetric distributions [20]. Because speech signals are known to be spherically invariant random processes in the frequency domain [21], such an assumption seems valid and also results in decent separation results. However, when compared to the frequency-domain ICA followed by perfect permutation correction, the separation performance of IVA using spherically symmetric joint densities is slightly inferior [19]. This suggests that such source priors do not exactly match the distribution of speech signals and that the IVA performance for speech separation can be improved by finding better dependency models [22,23].

We propose a new dependency model for IVA. The single and fully-connected clique is decomposed into many cliques of smaller sizes. A new objective function is derived to account for strong dependency inside the individual cliques and weak dependency across the cliques by retaining a considerable amount of overlap between adjacent cliques. The clique sizes are either fixed or determined by a mel-scale with its frequency index reversed; the latter was proven to be more robust to the increased number of cliques through simulated 2 × 2 speech separation experiments.

This article is organized as follows. Section 2 explains conventional IVA; Section 3 gives a detailed algorithm of the proposed method to contrast with IVA. Section 4 presents the results of the simulated speech separation experiments, and Section 5 summarizes the proposed method and its future extensions.

2 IVA

The key idea behind IVA is that all of the frequency components of a single source are regarded as a single vector, the components of which are dependent on one another. The independence between source vectors is approximated by a multivariate, joint probability density function (pdf) of the components from each source vector, and the joint pdf is maximized rather than the individual independencies between each frequency bin. The IVA model consists of a set of basic ICA models where the univariate sources across different dimensions have some dependency such that they can be grouped and aligned as a multidimensional variable.

Figure 1 illustrates a 2 × 2 IVA mixture model. Let the multidimensional source vector be $\mathbf{s}_i = [s_i^1, s_i^2, \dots, s_i^B]^T$ for $i = 1, 2$. Each component of $\mathbf{s}_1$ is linearly mixed with the component in the same dimension of $\mathbf{s}_2$ by $\mathbf{A}^b$ such that

$$\begin{aligned}\begin{bmatrix} x_1^b \\ x_2^b \end{bmatrix} &= \begin{bmatrix} a_{11}^b & a_{12}^b \\ a_{21}^b & a_{22}^b \end{bmatrix} \begin{bmatrix} s_1^b \\ s_2^b \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^b s_1^b + a_{12}^b s_2^b \\ a_{21}^b s_1^b + a_{22}^b s_2^b \end{bmatrix},\end{aligned} \qquad (5)$$

for $b = 1, \dots, B$. For microphone $j = 1, 2$, the observation vector is expressed as

$$\mathbf{x}_j = \begin{bmatrix} x_j^1 \\ x_j^2 \\ \vdots \\ x_j^B \end{bmatrix} = \begin{bmatrix} a_{j1}^1 s_1^1 + a_{j2}^1 s_2^1 \\ a_{j1}^2 s_1^2 + a_{j2}^2 s_2^1 \\ \vdots \\ a_{j1}^B s_1^B + a_{j2}^B s_2^B \end{bmatrix}. \qquad (6)$$

The mixing of the multivariate sources is dimensionally constrained so that a linear mixture model is formulated in each layer. The instantaneous ICA is extended to a formulation with multidimensional variables or vectors, where the mixing process is constrained to the sources on the same horizontal layer or on the same dimensions. The joint dependency within the dependent sources is modeled by a multidimensional pdf, and hence, correct permutation is achieved.

To derive the objective function of IVA, a single dimension of the estimated sources in Equation 4 is extracted, and a new vector is constructed by collecting the source coefficients of all the frequency bins. The source estimate $\mathbf{y}_i$ is expressed by the following matrix-vector multiplication:

$$\mathbf{y}_i = \begin{bmatrix} y_i^1 \\ y_i^2 \\ \vdots \\ y_i^B \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^N w_{ij}^1 x_j^1 \\ \sum_{j=1}^N w_{ij}^2 x_j^2 \\ \vdots \\ \sum_{j=1}^N w_{ij}^B x_j^B \end{bmatrix} \qquad (7)$$

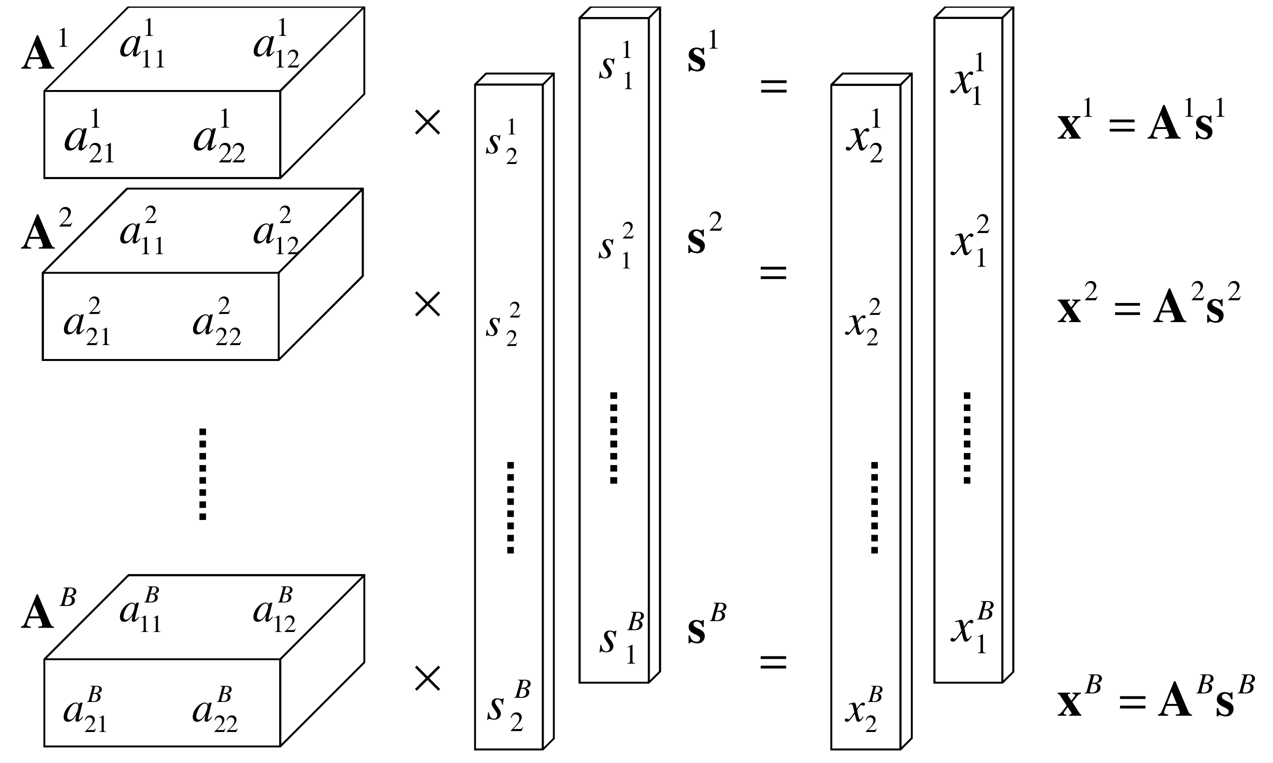

Figure 1 Mixture model of IVA.

$$= \begin{bmatrix} \mathbf{w}_i^1 \\ \mathbf{w}_i^2 \\ \vdots \\ \mathbf{w}_i^B \end{bmatrix} [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_N], \tag{8}$$

where $\mathbf{w}_i^b$ is the ith row of matrix $\mathbf{W}^b$ and $\mathbf{w}_{ij}^b$ is the jth element of $\mathbf{w}_i^b$. For a simple derivation of the IVA algorithm, we assume that y_i^b has a unit variance to eliminate the variance terms from the original IVA learning algorithm [19]. This can easily be achieved by scaling $\mathbf{w}_i^b$ appropriately such that

$$\mathbf{w}_i^b \leftarrow \mathbf{w}_i^b \Big/ \sqrt{E\left[|y_i^b|^2\right]}. \tag{9}$$

In resynthesis, the above normalization is reversed to restore the original scales. The likelihood of $\mathbf{y}_i$ is computed by the following multivariate pdf [19,20]:

$$p(\mathbf{y}_i) \propto \exp\left(-\sqrt{\sum_{b=1}^{B} |y_i^b|^2}\right). \tag{10}$$

The goal of IVA is optimizing $\{\mathbf{W}^1, \mathbf{W}^2, \dots, \mathbf{W}^B\}$ to maximize the independence among the separated sources, $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$, where the independence is approximated by the sum of the log likelihoods of the given data computed by Equation (10). The detailed learning algorithm can be found in [19,20].

Figure 2 illustrates the mixing assumption and how the IVA algorithm works. Two sources are mixed at different amounts in different frequency bins. To find $\mathbf{y}_1$ and $\mathbf{y}_2$ for the estimates of $\mathbf{s}_1$ and $\mathbf{s}_2$, IVA instead estimates the unmixing matrices to minimize the dependency between different sources while maintaining strong dependency across frequency bins. There is only a single dependency model in which all the frequency bins distinguished by their center frequencies are connected to one another: that is, the spherical dependency described by Equation (10).

3 Proposed dependency models for IVA

For real-sound sources, it is unreasonable for neighboring and distant frequency components to be assigned the same dependency because the dependency of neighboring frequency components is much stronger than that of distant frequency components. This section describes the proposed dependency models in which the single and fully connected statistical dependency of IVA is decomposed into several cliques whose sizes are set to be fixed or mel-scaled. The details of the proposed models are explained in this section.

3.1 Overlapped cliques of a fixed size

The statistical dependency between adjacent frequency components is much larger than that between distant components. For example, the dependency between y_i^b and y_i^{b+1} for an arbitrary b is much stronger than that between y_i^b and y_i^{b+k} when $k \gg 1$. We considered the difference in center frequencies of the STFT components in the proposed dependency model. As shown in Figure 3, the clique of the components of the estimated source vectors $\mathbf{y}_i$ was broken into several cliques in order to eliminate the direct dependency between distant frequency bins. This segmentation of the spherical model can be visualized as a chain of cliques [23]. The dependency among the source components propagates through chain-like overlaps of spherical dependencies such that the dependency between components weakens as the distance between them grows. The corresponding multivariate pdf is given in the following form:

$$p(\mathbf{y}_i) \propto \exp\left(-\sum_{c=1}^{C} \sqrt{\sum_{b=f_c}^{l_c} |y_i^b|^2}\right), \tag{11}$$

where C is the number of cliques, and f_c and l_c are the first and last indices, respectively, of clique c designed to satisfy the condition

$$f_c < l_{c-1}, \quad c = 2, 3, \dots, C, \tag{12}$$

so that the series of cliques have chained overlaps. With the proposed source prior in Equation (11), we derive a new learning algorithm to find a set of linear transformation matrices that make the components as statistically independent as possible, such that

$$\{\mathbf{W}^{b*}\} = \underset{\{\mathbf{W}^b\}}{\arg\max}\, \mathcal{L}(\{\mathbf{W}^b\}), \tag{13}$$

where the log likelihood function $\mathcal{L}$ is defined as

$$\begin{aligned} \mathcal{L}(\{\mathbf{W}^b\}) &\propto \log\left[\prod_b^B |\det \mathbf{W}^b| \cdot \prod_i^M p(\mathbf{y}_i)\right] \\ &= \sum_b^B \log|\det \mathbf{W}^b| + \sum_i^M \log p(\mathbf{y}_i) \\ &= \sum_b^B \log|\det \mathbf{W}^b| - \sum_i^M \sum_{c=1}^C \sqrt{\sum_{b=f_c}^{l_c} |y_i^b|}, \end{aligned} \tag{14}$$

where M is the number of sources defined in Equation (1). We apply the natural gradient learning rule [24] to $\mathbf{W}^b$ at each frequency bin b:

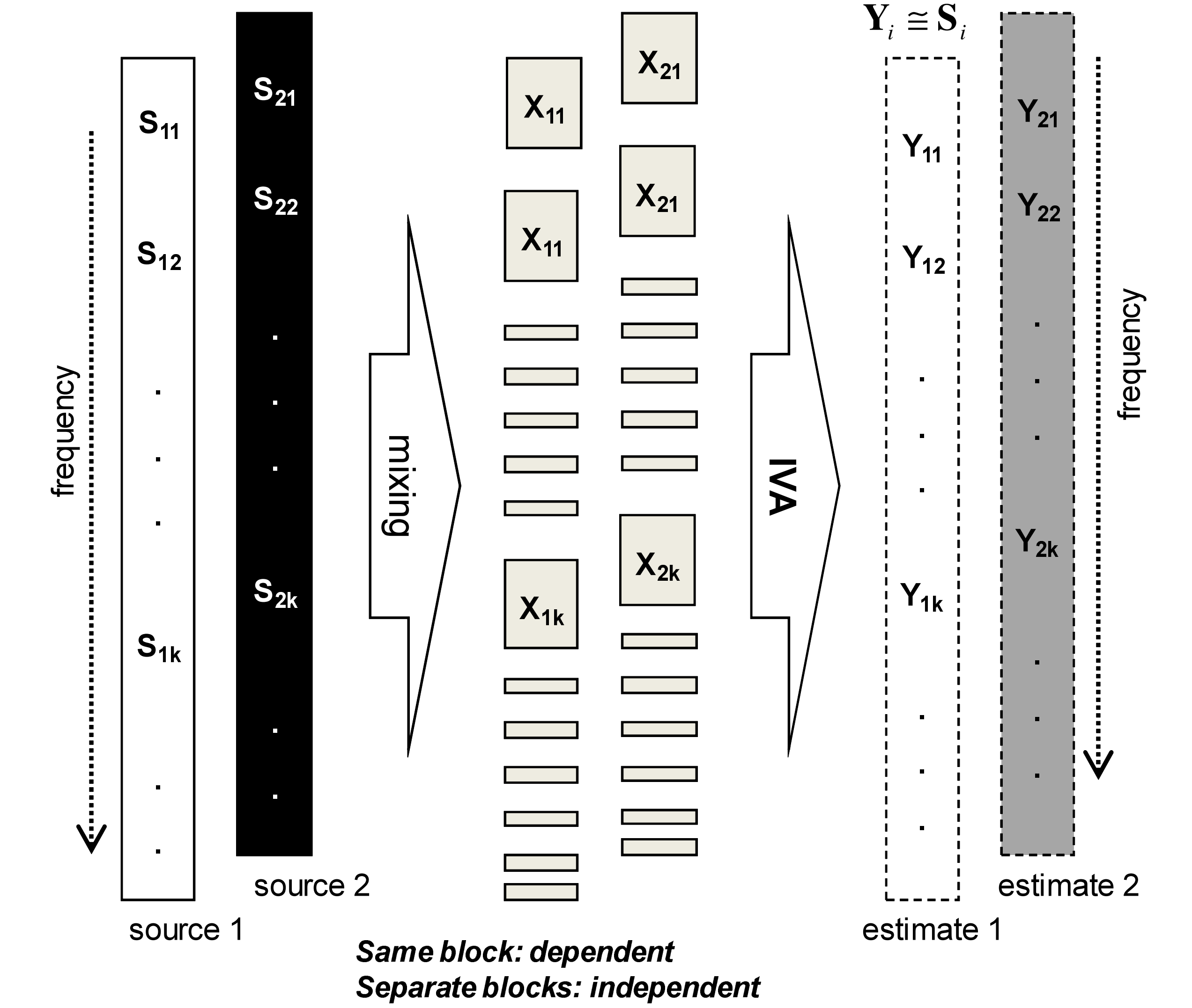

Figure 2 Mixing and separation models of conventional IVA.

$$\Delta \mathbf{W}^b \propto \left[\mathbf{I} - \varphi\left(\mathbf{y}_i^b\right)\mathbf{y}_i^b\right]\mathbf{W}^b, \tag{15}$$

where $\mathbf{I}$ is an $M \times M$ identity matrix, $(\cdot)^H$ is the Hermitian transpose operator, and $\phi(\mathbf{y}^b)$ is a vector function whose i^{th} element is

$$\begin{aligned}\left[\varphi(\mathbf{y}^b)\right]_i &= \frac{\partial \log p(y_i^b)}{\partial y_i^b} \\ &= \sum_{c \in \mathcal{S}_b} \frac{y_i^b}{\sqrt{\sum_{b=f_c}^{l_c} |y_i^b|^2}},\end{aligned} \tag{16}$$

where $\mathcal{S}_b$ is a set of cliques that includes bin b. At every adaptation step, $\mathbf{W}^b$ is constrained to be orthogonal by the following symmetric decorrelation scheme:

$$\mathbf{W}^b \leftarrow \left(\mathbf{W}^b(\mathbf{W}^b)^H\right)^{-\frac{1}{2}}\mathbf{W}^b, \quad b = 1, 2, \ldots, B. \tag{17}$$

At the end of the learning, the well-known minimal distortion principle [25] is applied to $\mathbf{W}^b$ by

$$\mathbf{W}^b \leftarrow \operatorname{diag}\left((\mathbf{W}^b)^{-1}\right)\mathbf{W}^b, \quad b = 1, 2, \ldots, B. \tag{18}$$

To select an appropriate set of cliques that is suited to our goal, we constructed a matrix of size $B \times B$ whose (i, j)th element is the correlation coefficient between bin i and bin j from a single source. Figure 4A-D shows the computed correlation coefficient matrices obtained from four different speech signals of two females and two males. In all four cases, a strong correlation was

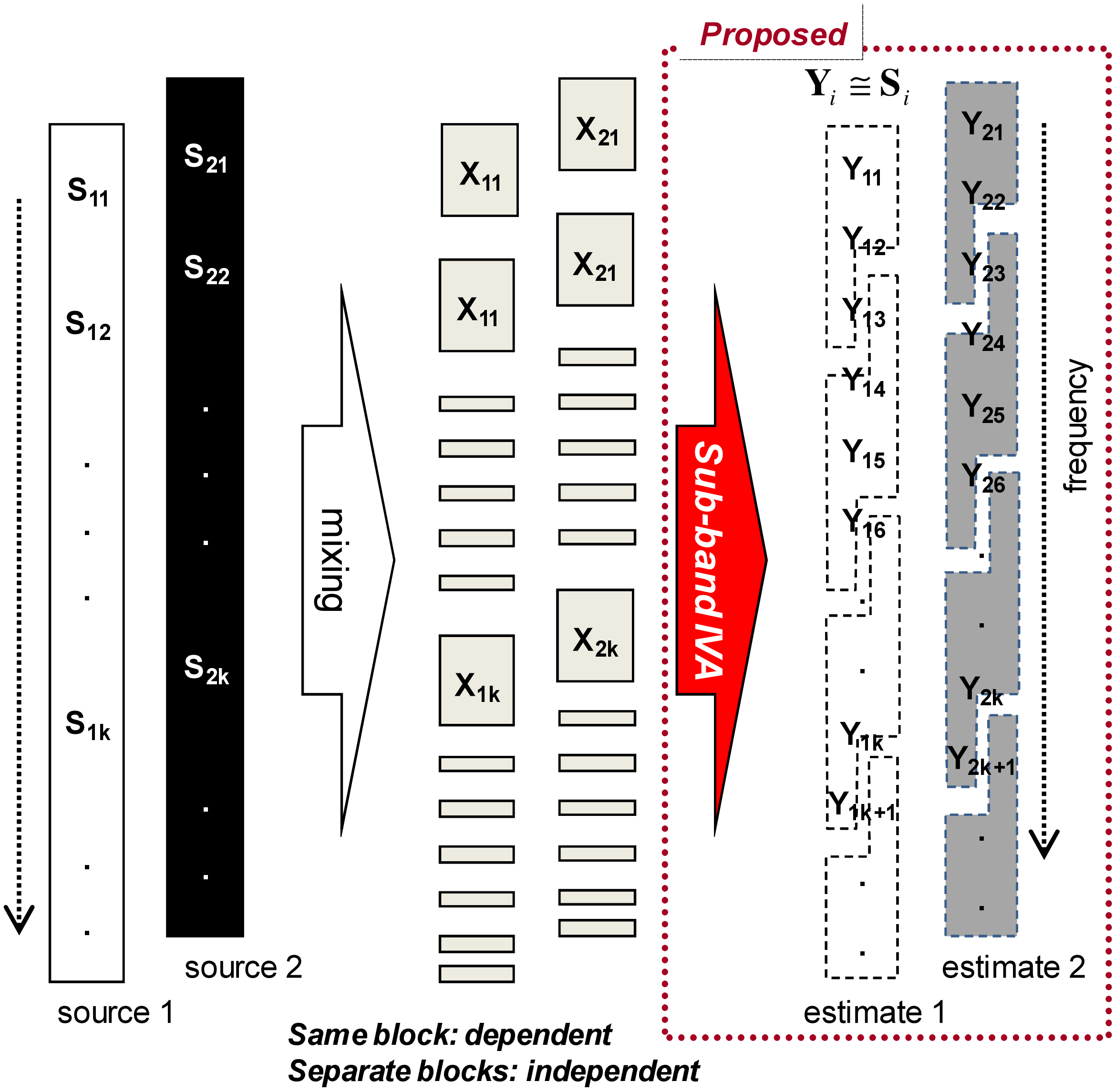

Figure 3 Illustration of the proposed dependency model.

observed around the diagonal with a positive slope because they were from closely located frequency pairs. The correlation decreased as it went off-diagonal. Although the low-frequency components had a widespread dependence over the 0-3 kHz region, it was much weaker than that along the positive sloping diagonal. All of the speech signals are from the TIMIT database, and the same observations held true for other speech signals as well. To consider strong correlations among neighboring frequency bins, we adopted a dependency graph consisting of several cliques of the same size and increasing center frequencies. Taking 1,024 frequency bins as an example, the beginning and ending indices of Equation (11) were $[f_1\ l_1]$ = [1 256], $[f_2\ l_2]$ = [2 257], $[f_3\ l_3]$ = [3 258], . . ., $[f_C\ l_C\ ]$ = [769 1024], where the number of frequency bins for each clique was fixed to 256. This simple dependency model using overlapped cliques is shown in Figure 5. All of the cliques were of the same size but with varying center frequencies.

3.2 Overlapped cliques of variable sizes

Figure 6 shows another model that reflects the spread dependence at low frequencies. The cliques have variable sizes based on the reversed mel-frequency scale. We adopted the mel-scale to prevent being biased to any specific cases; this scale has been proven to be efficient in numerous speech signal-processing applications

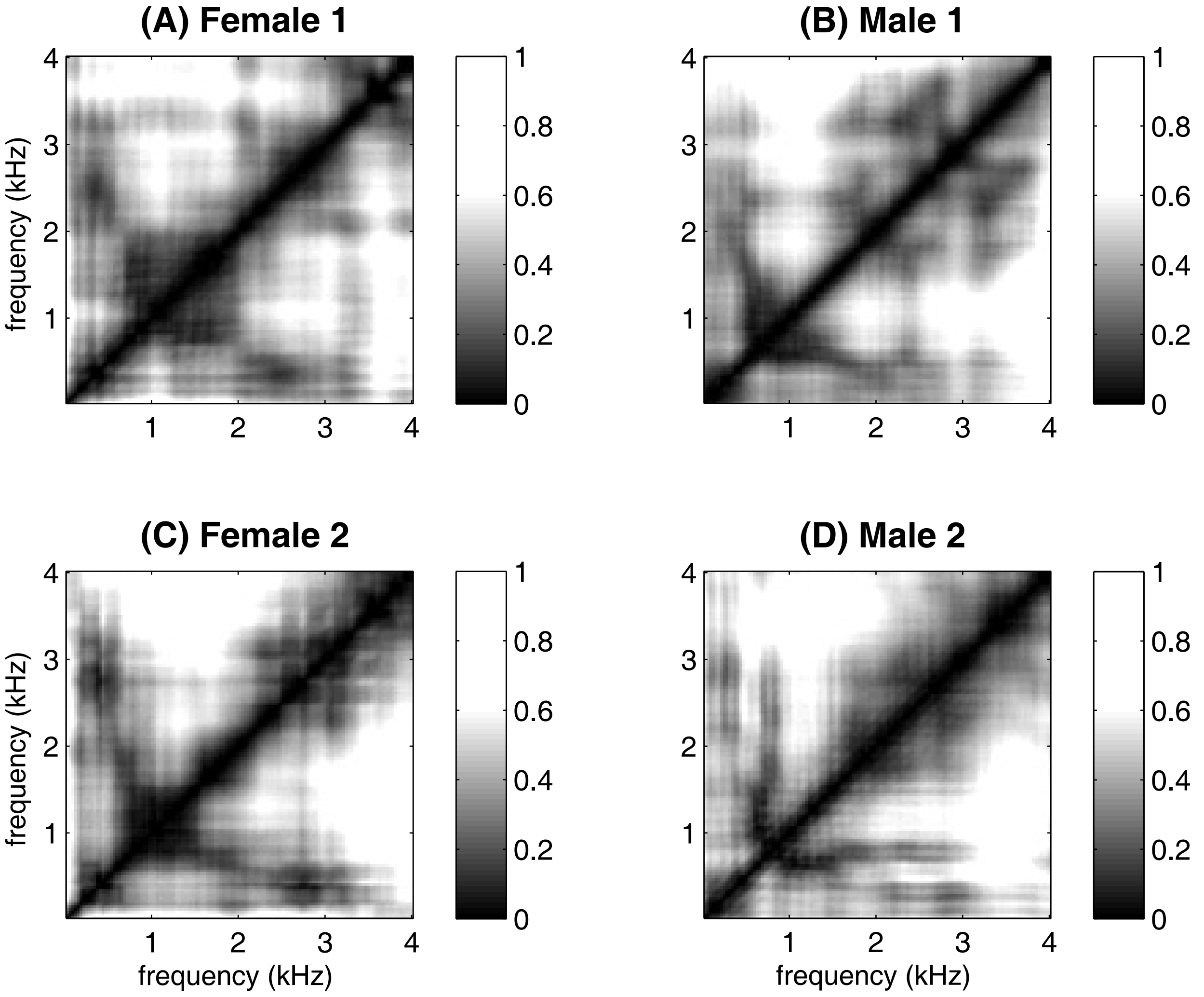

Figure 4 Correlation coefficient matrix of (A) female speech; (B) male speech; (C) another female; (D) another male. All speech signals are from the TIMIT database.

such as speech recognition and enhancement. General human speech is characterized by rapid changes occurring more often in the lower-frequency regions. Therefore, most auditory frequency scales, including the mel-scale, use a narrow bandwidth for the low-frequency region based on the observation that there is little dependence among neighboring frequencies [26]. In the high-frequency region, there is greater dependence among neighboring frequencies, so a relatively large bandwidth is used. However, in the proposed method, we set the sizes of the bands in the opposite fashion. We assigned larger clique sizes to low frequencies because they have less statistical dependence to one another, and smaller clique sizes to higher frequencies. Since the cliques play the role of joining the same source components distributed in different frequencies, a larger bandwidth is necessary to cover the weak and spread dependence in the low-frequency region. For higher frequencies, a smaller amount of overlap is enough because of the greater dependence among neighboring frequency components, as shown in Figure 4. The overlapped vertices between the adjacent cliques in the dependency graph enables collection of the same source components. Therefore, the clique size is determined by the reversed mel-scale, which is computed by

$$h\left(\omega_c\right) = A\left[\log_{10}\left(1 + \frac{\omega_c}{700}\right) - \log_{10}\left(1 + \frac{\omega_c - 1}{700}\right)\right], \quad (19)$$

where ω_c is the center frequency of clique c, A is a constant, and $h(\omega_c)$ is the bandwidth of clique c. The

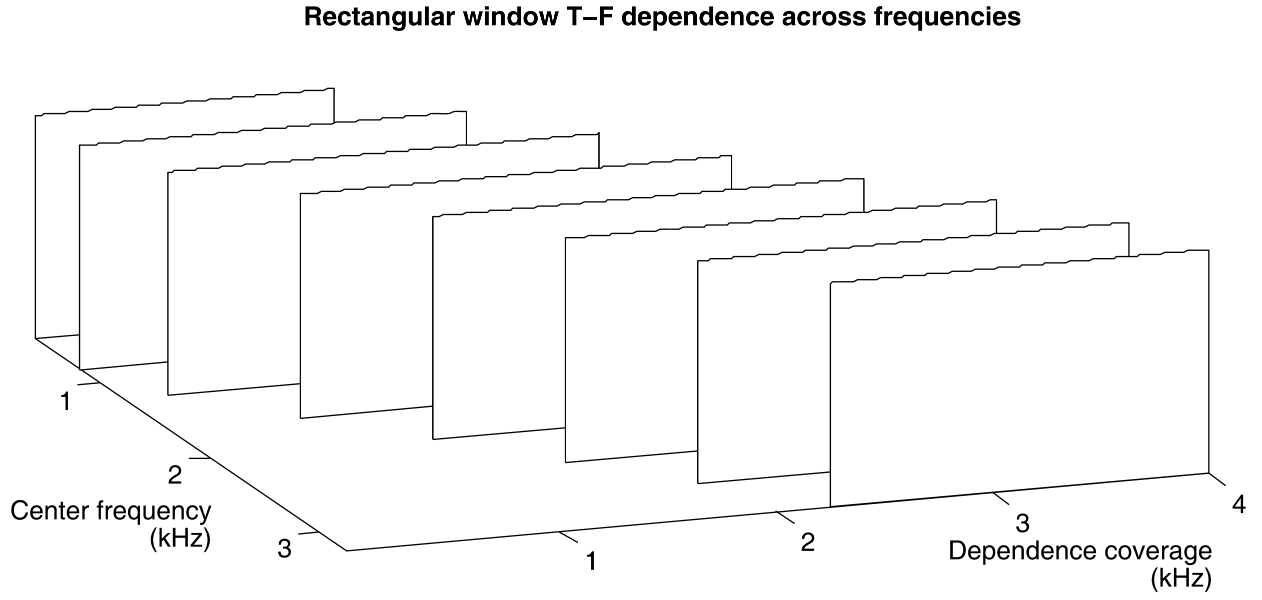
Rectangular window T−F dependence across frequencies
Center frequency (kHz)
1
2
3
Dependence coverage (kHz)
1
2
3
4

Figure 5 Dependency model of fixed clique size.

beginning and ending indices f_c and l_c in Equation (11) are then obtained by

$$f_c = \max\left(1, b_c - h\left(\omega_c\right)\right), \quad l_c = \min\left(B, b_c + h\left(\omega_c\right)\right), \quad (20)$$

where b_c is the center-bin number of clique c. The max and min operators ensure that the computed bin numbers are within a valid range.

4 Experiments

We compared the performance of the audio source separation using the proposed dependency models with that of the fully connected dependency model of the conventional IVA. Both methods were applied to multiple speech separation problems. The geometric configuration for the simulated room environments is shown in Figure 7. Various 2×2 cases were simulated by combining pairs of source locations from A to J. For example, experiment 1 was a combination of sound source 1 from location A and sound source 2 from location H, experiment 2 was a combination of sources from locations B and G, etc. We set the dimensions of the room to 7 m × 5 m × 2.75 m and the heights of all microphones and source locations to 1.5 m. The reverberation time was 100 ms, and the corresponding reflection coefficient was 0.57 for every wall, floor, and ceiling. Room

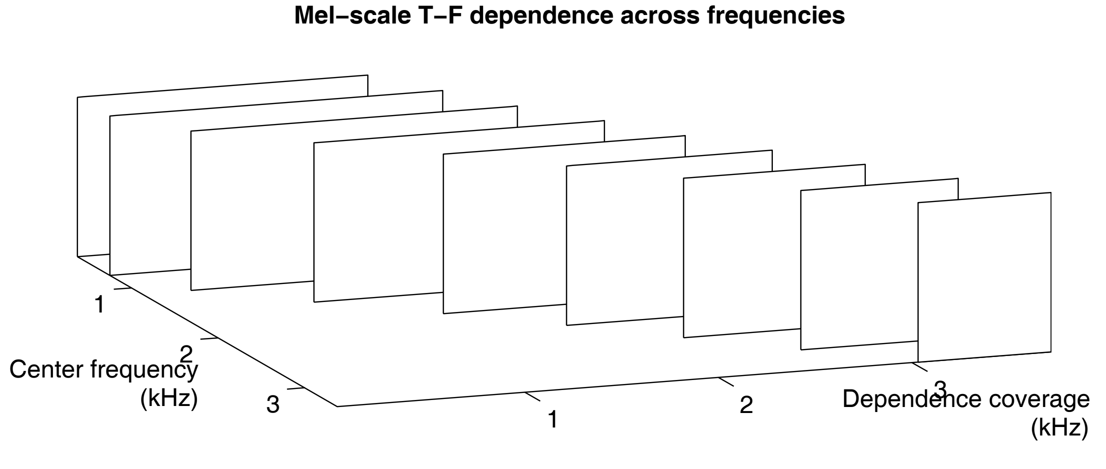
Mel−scale T−F dependence across frequencies
Center frequency (kHz)
1
2
3
Dependence coverage (kHz)
1
2
3

Figure 6 Dependency model of mel-scale clique sizes.

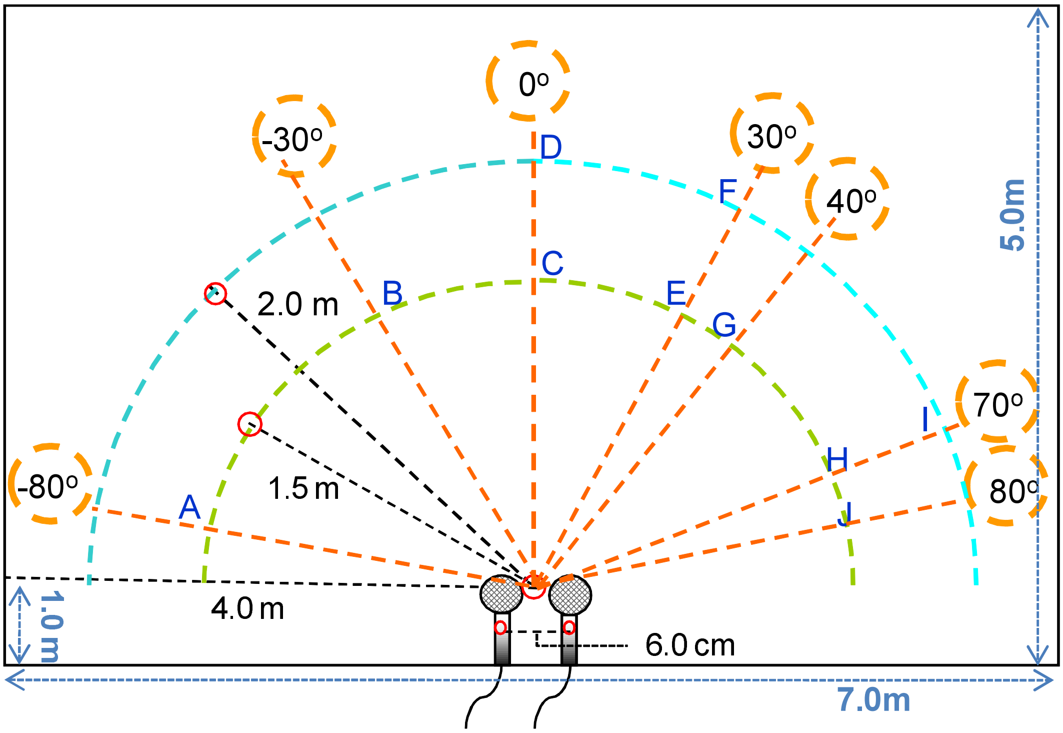

Figure 7 Geometric configuration of the simulated room environments.

impulse responses were obtained by an image method [1-3] using the above parameters. The impulse responses of the transfer functions from source locations A and H to the two microphones are shown in Figure 8. The peak location was not at the origin because the direct path had its own delay. The filter length was 100 ms, which was equivalent to 800 tabs at an 8-kHz sampling rate. The amplitude dropped rapidly because of the loss of energy due to the reflection.

Male and female speech signals chosen from the TIMIT database were synthetically convolved with the impulse responses corresponding to the locations of the

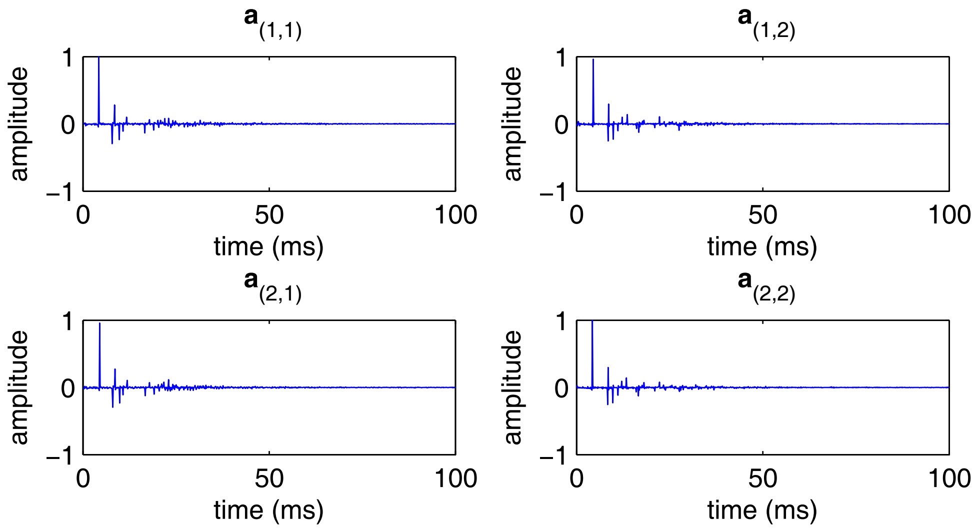

Figure 8 Impulse responses of the transfer function of a simulated room. The source locations are A and H.

sources and microphones in each experiment. When the algorithm was applied to source separation in the STFT domain, a 2048-point FFT, 2048-tab Hanning window, and shift size of 512 samples were used. The separation performance was measured in terms of the signal-to-interference ratio (SIR) which is defined as [19]:

$$\mathrm{SIR} = 10\log\left(\frac{\sum_{k,b}|\sum_{i} r^{b}_{iq(i)}s^{b}_{q(i)}[k]|^2}{\sum_{n,b}|\sum_{i\neq j} r^{b}_{iq(j)}s^{b}_{q(j)}[k]|^2}\right), \tag{21}$$

where $q(i)$ indicates the separated source index of the ith source and $r_{iq(j)}$ is the overall impulse response computed by $r_{iq(j)} = \sum_m w^{b}_{im}a^{b}_{mq(j)}$. In order to represent how close the estimated $\mathbf{W}^{b}_{i}$ was to the inverse of the mixing filters $\mathbf{A}^{b}_{j}$, the SIR numbers were measured in decibels, because the acoustic signal power ratio is in the log scale [26]. The higher SIR is, the closer the result is to perfect separation.

We compared the single clique model of IVA with the proposed multiple clique models. The multiple clique designs are shown in Figure 9. The numbers of cliques were 2, 4, 8, 12, and 16, and the overlap ratio between neighboring cliques was set to 50%. In A-E, the center frequencies were "linearly" increased, and the sizes were all fixed except for the first and last because they were located at opposite ends. For example, the four cliques in Figure 9B cover the frequency regions of 0-1.5, 0.5-2.5, 1.5-3.5, and 2.5-4 kHz. The neighboring cliques overlap by 50%, so the dependency is well propagated. In contrast, the center frequencies of F-J are on the "reversed mel-scale" in Equation (19): the clique sizes are inversely proportional to the rate of change in the mel-scale. The same four cliques in Figure 9G cover 0-2.2, 1.1-3.1, 2.4-3.7, and 3.3-4 kHz. Their actual bandwidths were 2.2, 2.0, 1.36, and 0.74 kHz, although the bandwidths computed by Equation (19) were 1.47, 1.02, 0.68, and 0.49 kHz. Because the first and last cliques had only one neighbor, their sizes were 1.5 times larger than the expected bandwidths, while the sizes of the second and the third cliques were twice as large to impose a 50% overlap with neighboring cliques.

The "CR" number in each of the clique designs in Figure 9 is the ratio of the sum of correlation coefficients enclosed by the union of all the cliques to the sum of the total correlation coefficients. It approaches unity as the enclosed region approaches the total area. The correlation map is identical to Figure 4A from the speech of female 1, who was one of the input sources of our experiments. The CR number does not account for the separation performance directly but roughly shows how well a clique design models the dependence of the frequency bins.

All of the separation performances were measured for their SIR and are summarized in Table 1. The first "IVA" row represents the SIR numbers obtained by the conventional IVA algorithm [19]. Rows labeled "LIN2," "LIN4," "LIN8," "LIN12," and "LIN16" are the results of the proposed models utilizing the clique designs in Figure 9A-E, and rows labeled "MEL2," ..., "MEL16" are the results with the clique designs in Figure 9F-J. Columns indicate various combinations of source locations, average SIR (denoted by "SIR") of the seven experiments, average number of iterations (denoted by "Iter.") for the solution to converge, and CR number of the corresponding clique design. The average SIRs that were

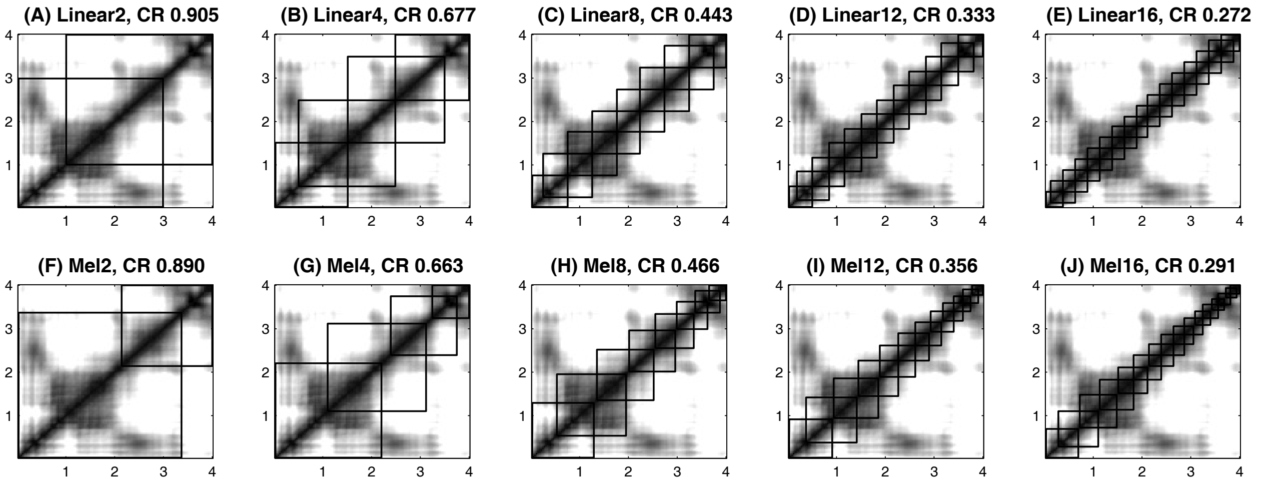

Figure 9 Various clique designs. **(A)-(E)** The center frequencies are linearly scaled, and the clique sizes are equal. **(F)-(J)** The center frequencies and clique sizes are on an inverse mel-scale. The "CR" values are the ratios of the sum of the correlation coefficients included in the cliques to the sum of all of the coefficients.

Table 1 Separation performances in SIR (dB)

Exp. number	1	2	3	4	5	6	7	Average		
Source location	**A,H**	**B,G**	**E,G**	**H,J**	**C,D**	**E,F**	**H,I**	**SIR**	**Iter.**	**CR**
IVA	16.5	17.5	16.6	12.0	15.5	15.2	15.0	15.5	936	1.000
LIN2	21.5	19.6	19.3	14.2	17.2	18.7	17.1	18.2	674	0.905
LIN4	22.0	19.6	19.3	14.7	17.4	19.2	18.4	18.7	397	0.677
LIN8	22.7	19.9	19.5	14.9	17.5	19.2	18.2	18.8	544	0.443
LIN12	7.3	18.8	9.0	5.8	17.6	19.6	18.8	13.9	468	0.333
LIN16	11.8	1.8	10.1	8.2	16.9	17.1	18.3	12.0	493	0.272
MEL2	19.4	19.0	18.5	13.5	16.7	16.5	15.8	17.1	825	0.890
MEL4	22.0	19.8	19.4	14.5	17.6	19.2	18.2	18.7	543	0.663
MEL8	22.3	19.8	19.3	14.7	17.3	19.1	18.6	18.7	408	0.466
MEL12	20.4	18.7	19.4	14.8	17.2	19.4	18.2	18.3	922	0.356
MEL16	20.9	19.0	18.6	14.9	17.5	19.2	18.2	18.3	1000	0.291

larger than 18 dB are boldfaced. Among the linear scales, the average SIRs of LIN4 and LIN8 were 18.7 and 18.8 dB, and the average numbers of iterations were 397 and 544, respectively. These indicate that LIN4 and LIN8 greatly improved both the separation performance and convergence speed compared to IVA. However, when the number of cliques was more than 8, SIR degraded rapidly (13.9 and 12.0 dB), and the separation performance were even poorer than those of IVA. Among mel-scales, the average SIRs of MEL4 and MEL8 were both 18.7 dB, and their numbers of iterations were 543 and 408, respectively, which were about the same as those of LIN4 and LIN8. The difference from the linear scales was when the number of cliques was more than 8: the separation performance measured by SIR did not degrade as badly as that of LIN12 and LIN16. However, many more number of iterations was required for both MEL12 and MEL16, implying that the broken dependency made the algorithm oscillate around the optimal solution. When comparing LIN and MEL, their best SIRs were almost the same, but the average iterations revealed that the mel-scales were more robust for large numbers of cliques. This can be explained by comparing the amount of correlation captured by the clique designs. Figure 9 shows that the CR numbers of MEL12 and MEL16 were 0.356 and 0.291, and those of LIN12 and LIN16 were 0.333 and 0.272, respectively. For 12 and 16 cliques, mel-scale designs had CR numbers larger by about 0.02 than the linear-scale designs. The difference mostly originated from the low-frequency region: the spread dependence observed at 1-2 kHz was better captured by the mel-scale cliques, and which in turn enabled correct source permutation. In summary, the proposed method was more effective than the original IVA in most clique configurations in terms of separation performance, and the mel-scale clique designs were better than the fixed-size designs in terms of stability.

5 Conclusions

The totally spherical dependency model of IVA was relaxed by the dependency models of chained cliques. The new clique designs are advantageous because the weak dependency among distant frequencies is modeled by indirect dependency propagation, which helps in finding a better local solution compared to the original IVA, where the same amount of dependency is assigned to any pair of frequency bins. In this article, two types of non-spherical models are proposed. The first uses the same number of frequency bins for all of the cliques, while the other varies the number of frequency bins in reversed mel-scales based on the measured correlation coefficients between different frequency bins. Both dependency models achieved higher source separation performance and faster convergence to correct solutions owing to more accurate modeling of the statistical dependency. For simulated mixtures of male and female speech signals, both models obtained the highest performance when the number of cliques was set to 4 or 8. When the clique size was fixed, the performance degraded drastically for more than eight cliques. However, when the clique size was determined by the mel-scales, the same level of performance was kept at the expense of convergence rate. This implies the presence of up to 16 independent units in speech signals along the mel-scale frequency axis. One of the ongoing research issues is finding more flexible dependency models, such as instantaneously varying the dependency graph based on the correlation coefficients measured from the input signals or on their harmonic structures. Another research issue is finding appropriate dependency models for natural sounds because the dependency among the frequency components may not be related to the mel-scale.

Acknowledgements
This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0025642), and by the Ministry of Knowledge Economy, Korea (2008-S-019-02, Development of Portable Korean-English Automatic Speech Translation Technology).

Author details
[1]Hamilton Glaucoma Center, University of California, San Diego, CA, USA
[2]Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea

Competing interests
The authors declare that they have no competing interests.

Received: 15 June 2011 Accepted: 23 May 2012 Published: 23 May 2012

References

1. RB Stephens, AE Bate, *Acoustics and Vibrational Physics*, (Edward Arnold Publishers, London, 1966)

2. JB Allen, DA Berkley, Image method for efficiently simulating small room acoustics. J Acoust Soc Am. **65**, 943–950 (1979). doi:10.1121/1.382599
3. WG Gardner, The virtual acoustic room. Master's thesis, MIT (1992)
4. AJ Bell, TJ Sejnowski, An information maximization approach to blind separation and blind deconvolution. Neural Comput. **7**(6), 1129–1159 (1995). doi:10.1162/neco.1995.7.6.1129
5. D Yellin, E Weinstein, Multichannel signal separation: methods and analysis. IEEE Trans Signal Process. **44**, 106–118 (1996). doi:10.1109/78.482016
6. K Torkkola, Blind separation of convolved sources based on information maximization, in *Proc IEEE Int Workshop on Neural Networks for Signal Processing*, Kyoto, Japan, 423–432 (1996)
7. R Lambert, Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures, PhD thesis, (University of Southern California, 1996)
8. TW Lee, AJ Bell, R Lambert, Blind separation of delayed and convolved sources. Adv Neural Inf Process Syst. **9**, 758–764 (1997)
9. A Hyvärinen, E Oja, *Independent Component Analysis*, (John Wiley and Sons, New York, 2002)
10. P Smaragdis, Blind separation of convolved mixtures in the frequency domain. Neurocomputing. **22**, 21–34 (1998). doi:10.1016/S0925-2312(98)00047-2
11. L Parra, C Spence, Convolutive blind separation of non-stationary sources. IEEE Trans Speech Audio Process. **8**(3), 320–327 (2000). doi:10.1109/89.841214
12. F Asano, S Ikeda, M Ogawa, H Asoh, N Kitawaki, A combined approach of array processing and independent component analysis for blind separation of acoustic signals. in *Proc IEEE Int Conf on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah. **5**, 2729–2732 (2001)
13. J Anemueller, B Kollmeier, Amplitude modulation decorrelation for convolutive blind source separation, in *Proc Int Conf on Independent Component Analysis and Blind Source Separation*, Helsinki, Finland, 215–220 (2000)
14. N Murata, S Ikeda, A Ziehe, An approach to blind source separation based on temporal structure of speech signals. Neurocomputing. **41**, 1–24 (2001). doi:10.1016/S0925-2312(00)00345-3
15. J Anemueller, TJ Sejnowski, S Makeig, Complex independent component analysis of frequency-domain electroencephalographic data. Neural Netw. **16**(9), 1311–1323 (2003). doi:10.1016/j.neunet.2003.08.003
16. A Hiroe, Solution of permutation problem in frequency domain ICA, using multivariate probability density functions. Lecture Notes in Computer Science. **3889**, 601–608 (2006). doi:10.1007/11679363_75
17. I Lee, T Kim, TW Lee, Complex FastIVA: a robust maximum likelihood approach of MICA for convolutive BSS. Lecture Notes in Computer Science. **3889**, 625–632 (2006). doi:10.1007/11679363_78
18. I Lee, T Kim, TW Lee, in *Independent Vector Analysis for Convolutive Blind Speech Separation*, Chap 6. (Springer, New York, 2007), pp. 169–192
19. T Kim, H Attias, SY Lee, TW Lee, Blind source separation exploiting higher-order frequency dependencies. IEEE Trans Audio Speech Lang Process. **15**, 70–79 (2007)
20. I Lee, TW Lee, On the assumption of spherical symmetry and sparseness for the frequency-domain speech model. IEEE Trans Speech Audio Lang Process. **15**(5), 1521–1528 (2007)
21. H Brehm, W Stammler, Description and generation of spherically invariant speech-model signals. Signal Process. **12**(2), 119–141 (1987). doi:10.1016/0165-1684(87)90001-6
22. I Lee, GJ Jang, TW Lee, Independent vector analysis using densities represented by chain-like overlapped cliques in graphical models for separation of convolutedly mixed signals. Electron Lett. **45**(13), 710–711 (2009). doi:10.1049/el.2009.0945
23. GJ Jang, IT Lee, TW Lee, Independent vector analysis using non-spherical joint densities for the separation of speech signals, in *Proc IEEE Int Conf on Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, **2**, 629–632 (2007)
24. SI Amari, A Cichocki, HH Yang, A new learning algorithm for blind signal separation. Adv Neural Inf Process Syst. **8**, 757–763 (1996)
25. K Matsuoka, S Nakashima, Minimal distortion principle for blind source separation, in *Proc Int Conf on Independent Component Analysis and Blind Source Separation*, San Diego, California, 722–727 (2001)
26. D O'Shaughnessy, *Speech Communication: Human and Machine*, (Addison-Wesley, New York, 1987)

doi:10.1186/1687-6180-2012-113
Cite this article as: Lee and Jang: Independent vector analysis based on overlapped cliques of variable width for frequency-domain blind signal separation. *EURASIP Journal on Advances in Signal Processing* 2012 2012:113.