High-speed separation of anechoic mixtures of speech signals using fusion of iterative and closed-forms separation method

Mahdi MAHDIKHANI*, Mohammad Hossein KAHAEI
School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

Received: 21.02.2013 • Accepted/Published Online: 26.06.2013 • Printed: 31.12.2015

Abstract: We introduce the fusion of iterative and closed-forms separation (FICS) method for high-speed separation of anechoic mixed speech signals. This method is performed in two stages: 1) iterative-form separation and 2) closed-form separation. This algorithm significantly improves the separation quality simply due to incorporating only some specific frequency bins into computations. We apply the FICS method to the frequency-domain independent component analysis (ICA) to evaluate its performance in increasing the signal separation speed for anechoic mixtures. Simulation results show that for speech signals and anechoic conditions, the proposed algorithm is on average 65 times faster than ICA while preserving the separation quality. It also outperforms FastICA, JADE, and SOBI in terms of separation quality and speed.

Key words: Blind speech separation, independent component analysis, mixed speech signals

1. Introduction

Blind source separation (BSS) is used for separating different mixed signals received by sensors [1]. In such problems, the separation procedure is performed in the time or frequency domain [2] with a little knowledge about the mixing model and properties of source signals such as independency, sparsity, etc. [3]. Time-domain methods are computationally a burden due to assuming a convolutive mixture model on the contrary to frequency-domain methods, which are simpler due to converting a convolutive model into an instantaneous one. In turn, the latter methods involve the local permutation problem, which reduces the separation quality [4].

A number of BSS algorithms are designed in the time or frequency domain using the independent component analysis (ICA) algorithm [5] to address the problem of instantaneous blind separation [6,7]. In general, ICA algorithms are based on two approaches: 1) minimization of mutual information [8,9] and 2) maximization of non-Gaussianity [10,11].

Conventional ICA is a maximum likelihood algorithm [12,13] based on a minimization of mutual information approach [14–17] and using the natural gradient method [18]. Another algorithm presented in [19,20] is the fast fixed point (FastICA) algorithm for separation of complex valued, linearly mixed independent source signals. This algorithm is based on maximization of the non-Gaussianity approach [20], which is computationally efficient and a robust. In the second-order blind identification (SOBI) algorithm [21,22], the second-order cumulants are used to diagonalize covariance matrices, while in the joint approximate diagonalization of eigen-

*Correspondence: m_mehdikhani2000@yahoo.com
matrices (JADE) algorithm [23] the fourth-order cumulants [24] are also incorporated in order to perform a joint diagonalization of the cumulant matrices using the Jacobi technique. Cardoso et al. [25] demonstrated that JADE is equivalent to minimization of mutual information approach. The cost function used in [25] is effectively a measure of mutual information between the cross-cumulants of signals. Contrary to the maximum likelihood ICA, FastICA, SOBI, and JADE separate the signals faster at the expense of losing the separation quality [19].

In this paper, the maximum likelihood ICA, or simply the ICA, is selected as a reference for comparison of separation quality of mixed speech signals with an anechoic mixing model in the frequency domain [26]. A drawback of the ICA is its low separation speed, which can be controlled by increasing the rate of the step-size parameter [27,28]. However, this may lead to converging of the ICA by some undesired oscillations and thus reduction of the separation quality. In [29], we presented the detection and removing of learning curve oscillations (DR-LCO) method to speed up ICA by detecting and removing such oscillations while preserving the separation quality, leading to introducing the speeded-up ICA (SICA) algorithm. Although SICA is at least 3 times faster than ICA, still this result can be further improved. In [30], we presented the variable situated matrix (VSM) technique to speed up ICA without loss of quality, leading to introduction of the VICA algorithm. In the VICA algorithm, in all learning steps of each frequency bin the separation quality of the separating matrix is compared with our defined situated matrix and the best one is selected as an initial separating matrix in the next learning step. On average, the VICA algorithm is about 6 times faster than the ICA while the quality of separated signals remains almost unchanged or becomes slightly better.

In this work, we introduce the fusion of iterative and closed-forms separation (FICS) method to speed up separation of mixed speech signals. In this method, the separation procedure is first performed iteratively for some individual frequency bins and the attenuation and delay matrices are estimated. Next, the final estimated attenuation and delay matrices are applied to other frequency bins to separate the signals in closed form. The FICS method can be applied to any frequency-domain separating algorithms. However, in this paper, we apply the FICS method to the VICA to introduce the new FVICA (FICS+VSM+ICA) algorithm. Simulation results show that the FVICA algorithm is much faster than the ICA, while its separation quality is almost comparable. The FVICA algorithm also outperforms the well-known FastICA, JADE, and SOBI in terms of separation quality and speed.

The paper is organized as follows. In Sections 2 and 3, the mixture model and ICA are introduced, respectively. The FICS method is studied in Section 4. In Section 5, simulation results are presented, and Section 6 concludes the paper.

2. Mixture model
We intend to separate the mixed signals received from $N$ sources by a uniform linear array (ULA) with $M$ sensors located linearly with an equal space of $\delta$ as shown in Figure 1. In an ordinary ICA, we consider the determined case with $M = N$ while the first sensor is considered as the reference. In the overdetermined case in which we have $N < M$, the problem is easily reduced to a determined case by selecting $N$ sensors, or applying some more sophisticated preprocessing techniques like the principle component analysis (PCA) algorithm. For an anechoic mixture model, received signals are defined as [31]:

$$x_m(t) = \sum_{n=1}^{N} a_{mn}s_n(t - d_{mn}), \quad 1 \leq m \leq M, \tag{1}$$
where \( x_m(t) \) is the \( m_{th} \) sensor signal and \( s_n(t) \) is the \( n_{th} \) speech source signal. Also, \( a_{mn} \) and \( d_{mn} \) respectively show the attenuation and time delay of the \( n_{th} \) source signal with respect to the \( m_{th} \) sensor (see Appendix 1, on the journal’s website). Accordingly, in matrix form, the attenuation and delay matrices are defined as \( A = [a_{mn}], \quad D = [d_{mn}], \quad 1 \leq m, n \leq M \).

![Figure 1. A ULA with \( M \) sensors receiving signals from the \( n_{th} \) source.](image)

To separate the signals in the time-frequency domain, the STFT of Eq. (1) is expressed as:

\[
\begin{bmatrix}
X^r_1 \\
\vdots \\
X^r_m \\
\vdots \\
X^r_M
\end{bmatrix}
= 
\begin{bmatrix}
\psi^1_1 & \ldots & \psi^1_n & \ldots & \psi^1_N \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\psi^m_1 & \ldots & \psi^m_n & \ldots & \psi^m_N \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\psi^M_1 & \ldots & \psi^M_n & \ldots & \psi^M_N
\end{bmatrix}
\begin{bmatrix}
S^r_1 \\
\vdots \\
S^r_n \\
\vdots \\
S^r_N
\end{bmatrix},
\]

(2)

where

\[
\psi^m_n = a_{mn}e^{-j2\pi r\Delta f d_{mn}}
\]

(3)

and \( \Delta f \) denotes the frequency resolution. Also, \( X^r_m \) and \( S^r_n \) respectively show the \((r,l)_{th}\) component of \( R \times L \) matrices \( X_m \) and \( S_n \), which are the STFT of the \( m_{th} \) sensor and \( n_{th} \) source signals. \( R \) and \( L \) show the number of frequency bins and time windows, respectively. For simplicity, Eq. (2) is shown in matrix form as

\[
y_{r,l} = \Psi_r q_{r,l}.
\]

(4)

where \( y_{r,l} \) and \( q_{r,l} \) are respectively defined as

\[
y_{r,l} = \begin{bmatrix}
X^r_1 \\
\vdots \\
X^r_m \\
\vdots \\
X^r_M
\end{bmatrix}^T,
\]

(5)

\[
q_{r,l} = \begin{bmatrix}
S^r_1 \\
\vdots \\
S^r_n \\
\vdots \\
S^r_N
\end{bmatrix}^T,
\]

(6)

and \( \Psi_r = [\psi^{mn}_r] \) is an \( M \times N \) mixing matrix [32].
3. An overview of the ICA algorithm

In the maximum likelihood ICA, the likelihood function defined as

$$\xi(W_r | y_r,t) = \prod_{l=1}^{L} p(y_{r,t}|W_r), \quad (7)$$

maximized wherein $y_{r,1}, \ldots, y_{r,t}, \ldots, y_{r,L}$ are the observation vectors and $p(.)$ is the frequency domain density function. We also have

$$\hat{q}_{r,t} = W_r y_{r,t}, \quad (8)$$

where

$$\hat{q}_{r,t} = [ \hat{S}_{r1}^l \ldots \hat{S}_{rn}^l \ldots \hat{S}_{rN}^l ]^T, \quad (9)$$

with $\hat{S}_{rn}^l$ being the estimate of $S_{rn}^l$ and $W_r$ showing the separating matrix of the $r_{th}$ frequency bin. Then $p(y_{r,t}|W_r)$ is given by

$$p(\hat{q}_{r,t}) = \frac{1}{|\text{det}(W_r)|} p(y_{r,t}) \Leftrightarrow p(y_{r,t}|W_r) = |\text{det}(W_r)| p(\hat{q}_{r,t}). \quad (10)$$

Since the separated signals, $\hat{S}_{rn}^l$, are assumed to be independent of each other, we can write

$$p(\hat{q}_{r,t}) = \prod_{n=1}^{N} p(\hat{S}_{rn}^l). \quad (11)$$

By substituting Eq. (11) into Eq. (10) and then into Eq. (7), the log likelihood function is obtained as

$$\log \xi(W_r | y_r,t) = L \log |\text{det}(W_r)| + \sum_{l=1}^{L} \sum_{n=1}^{N} \log p(\hat{S}_{rn}^l), \quad (12)$$

where $p(\hat{S}_{rn}^l) = \exp(-\sqrt{\sum_{n=1}^{N}(\hat{S}_{rn}^l)^2 + \xi})$ with $\zeta = 0.1$ and $\xi = 1$ for speech signals [18]. Accordingly, in a simpler form, we consider the following cost function:

$$\zeta(W_r) = \log |\text{det}(W_r)| + \frac{1}{L} \sum_{l=1}^{L} \sum_{n=1}^{N} \log p(\hat{S}_{rn}^l). \quad (13)$$

Maximization is then performed using an iterative update equation and a small step-size parameter $\eta_0$ as

$$W_r \leftarrow W_r + \eta_0 \frac{\partial \zeta(W_r)}{\partial (W_r)^{v}}. \quad (14)$$

After some mathematical manipulation and using the natural gradient, the iterative learning algorithm of the ICA for the $v_{th}$ iteration is obtained as [18]:

$$W_r^{v+1} = W_r^v + \eta_0 \left[ I_M - \frac{1}{L} \Phi_r^v (\hat{Q}_r^v)^H \right] W_r^v, \quad (15)$$

2046
where $W^v_r$ is the $M \times M$ separating matrix of the $v_{th}$ learning iteration and $r_{th}$ frequency bin and $I_M$ and $H$ show the identity matrix and Hermitian operator, respectively. $\hat{Q}^v_r$ is a resulting $M \times L$ matrix of separated signals defined as $\hat{Q}^v_r = W^v_r Y_r$, where the $M \times L$ matrix $Y_r$ is given by

$$Y_r = \begin{bmatrix} y_{r,1} & \cdots & y_{r,l} & \cdots & y_{r,L} \end{bmatrix}$$

and $y_{r,l}$ was expressed in Eq. (5). $\Phi^v_r$ is defined as

$$\Phi^v_r = \begin{bmatrix} \phi \left( \hat{q}_{r,v}^{11} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{1L} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{ML} \right) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi \left( \hat{q}_{r,v}^{m1} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{ml} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{mL} \right) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \phi \left( \hat{q}_{r,v}^{M1} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{Ml} \right) & \cdots & \phi \left( \hat{q}_{r,v}^{ML} \right) \end{bmatrix}$$

where $\hat{q}_{r,v}^{ml}$ is the $(m,l)_{th}$ component of the $M \times L$ matrix $\hat{Q}^v_r$ and

$$\phi \left( \hat{q}_{r,v}^{ml} \right) = \frac{\hat{q}_{r,v}^{ml}}{2\sqrt{\left| \hat{q}_{r,v}^{ml} \right|^2 + \zeta}}.$$ 

After convergence of Eq. (15), the final separated signals matrix of the $r_{th}$ frequency bin, $W_r$, is obtained as $\hat{Q} = W_r Y_r$ where

$$W_r \simeq (\Psi_r)^{-1},$$

and $\hat{Q} = \begin{bmatrix} \hat{q}_{r,1} & \cdots & \hat{q}_{r,l} & \cdots & \hat{q}_{r,L} \end{bmatrix}$ is a $M \times L$ matrix and $\hat{q}_{r,l}$ was expressed in Eq. (9). By applying Eq. (15) to all the frequency bins and solving the permutation problem based on [4], the final separated signals are obtained in the time-frequency domain. Then we only need to transform the separated signals to the time domain using the ISTFT.

4. Introducing the FICS method

As known, a great amount of energy of mixed speech signals is concentrated in a few frequency bins in which we would expect to obtain a better separation quality and more accurate estimates of the mixing, attenuation, and delay matrices, $\Psi_r$, $A$, and $D$, because in these more energetic bins the values of the signal to noise ratio (SNR) is higher than that of the less energetic ones.

To inspect the separation quality, the measuring term, $20 \log_{10} \left| W_r - (\Psi_r)^{-1} \right|$, is plotted versus the frequency bin number in Figure 2 for $N = 2, 3, \text{and} \ 4$ and $r = 1$ to 256. Note that $\left\| . \right\|$ denotes the norm operator and gives the largest singular value of a matrix. Clearly, a lower value of this term in the $r_{th}$ bin shows the greater similarity of $W_r$ to $(\Psi_r)^{-1}$, presenting a more accurate separating matrix and better separation quality of speech signals. As a result, more accurate estimates of matrices $A$ and $D$ shown by $\hat{A}$ and $\hat{D}$ are obtained by applying the separating algorithm in a few frequency bins only.
On the other hand, the components of $A$ and $D$, shown by $a_{mn}$ and $d_{mn}$, are only dependent on relative positions of sources and sensors, which are assumed to be fixed and independent of frequency (see Appendix 1, on the journal’s website). Therefore, the $\hat{A}$ and $\hat{D}$ matrices can be used for other frequency bins to separate the signals in closed form.

The above fact motivates us to develop the FICS method to achieve a high-speed separation of mixed speech signals without loss of quality. In this method, unlike conventional frequency-domain separating algorithms, frequency bins are divided into two sets: selected frequency bins (SFBs) and remaining frequency bins (RFBs) sets with different separating procedures. In each bin of the SFB set, mixed signals are separated iteratively and attenuation and delay matrices are estimated. The final estimated $\hat{A}$ and $\hat{D}$ are then computed by averaging over all attenuation and delay matrices of all SFB bins, respectively. Accordingly, this stage is referred to as the iterative-form separation (IS). Having estimated $\hat{A}$ and $\hat{D}$, the separating matrix of each frequency bin of the RFB set is formed to carry out a closed-form separation (CS). As the main goal of this work, by applying a closed-form separation instead of an iterative one, a very large number of computations are avoided, leading to a very fast signal separation. Although the performance of the FICS method is dependent on both the IS and CS stages (related to SFB and RFB sets), not just one of them, as a rule of thumb, the separation quality and separation speed of the proposed FICS method are generally due to the IS and the CS stage, respectively. In the SFB set the most energy of the speech signals is concentrated and in the RFB set the closed-form separation is performed instead of the iterative one.

In general, the FICS method can be applied to any frequency-domain separating algorithm as an auxiliary module. Here we apply the FICS to the VICA [30] to introduce the FVICA (FICS+VSM+ICA) algorithm and investigate its separation speed and quality. Signal to distortion ratio (SDR), signal to interference ratio (SIR), and perceptual estimation of speech quality (PESQ) are calculated for measuring the separation quality (see Appendix 2, on the journal’s website).

The FVICA is detailed as follows for a given SFB set consisting of $K$ bins from $g$ to $g + K - 1$ with $1 \leq g \leq R$ and $1 \leq K \leq R + 1 - g$. However, it is vital to mention that although the FVICA can work with any SFB sets, in Section 5 we will show that for separation of mixed speech signals we only use specific sets of bins mentioned as the best SFB (bSFB) sets for which the FVICA yields the best performance in terms of
separation quality and speed. The procedure of applying the FVICA for mixed speech signals is performed in the two following stages.

**Stage 1: Iterative-form separation (IS)**

The procedure of Stage 1 is itemized as follows and shown in Figure 3.

1. Apply the frequency-domain VICA to each bin of the SFB set to iteratively obtain $W_r$ using Eq. (15) and calculate $\hat{Q}_r$, $g \leq r \leq g + K - 1$. Then $\hat{S}^r_n$, $g \leq r \leq g + K - 1$ are computed for the $n_{th}$ source. As a result, separation is performed for the $K$ bins shown in blue in Figure 4.

2. Estimate the mixing matrices $\hat{\Psi}_r = (W_r)^{-1}$, $g \leq r \leq g + K - 1$ from Eq. (19).

3. Estimate the components of attenuation and delay matrices $\hat{a}_{mn} = |\hat{\psi}_{r}^{mn}|$ and $\hat{d}_{mn} = \left(-\frac{1}{2\pi r\Delta f}\right) \angle(\hat{\psi}_{r}^{mn})$ from Eq. (3) in each bin of the SFB set for $1 \leq m, n \leq M$ where $|.|$ and $\angle(.)$ are amplitude and phase operators, respectively.

Figure 3. Procedure of Stage 1 of the FVICA algorithm.
4. Compute $\hat{a}_{mn}$ and $\hat{d}_{mn}$ by averaging $\tilde{a}_{mn}$ and $\tilde{d}_{mn}$ over $K$ bins of the SFB set, respectively.

5. Consist the attenuation and delay matrices $\hat{\mathbf{A}} = [\hat{a}_{mn}]$ and $\hat{\mathbf{D}} = [\hat{d}_{mn}]$.

By estimating $\hat{\mathbf{A}}$ and $\hat{\mathbf{D}}$, Stage 2 is started.

**Stage 2: Closed-form separation (CS)**

After estimating $\hat{\mathbf{A}}$ and $\hat{\mathbf{D}}$, the RFB set including $R - K$ bins are separated in closed form as itemized below.

1. Form the mixing matrix $\hat{\mathbf{S}}_r$ by components $\hat{s}^{\text{sr}}_{mn} = \hat{a}_{mn} e^{-j2\pi r \Delta f \hat{d}_{mn}}, \{1 \leq r \leq g - 1\} \cup \{g + K \leq r \leq R\}$.

2. Obtain $\mathbf{W}_r$ using Eq. (19) and calculate $\hat{\mathbf{Q}}_r$ and $\hat{S}^{\text{sr}}_r$, $\{1 \leq r \leq g - 1\} \cup \{g + K \leq r \leq R\}$ for the $n_{th}$ source. Therefore, separation is performed for the remaining $R - K$ bins shown in pink in Figure 4.

After obtaining the separated signals $\hat{S}^{\text{sr}}_n$ from Stages 1 and 2, we apply the ISTFT to estimate the whole separated speech signals $\hat{s}_n(t)$, $1 \leq n \leq N$ in the time domain.

Implementation of this stage is very fast because an iterative separation is replaced by a closed-form one.

**5. Simulation results**

Simulation results are expressed in two parts: finding the $b$SFB sets and comparing the FVICA performance for speech signals separation compared to other algorithms. In all the experiments, sources generate speech signals with the duration of 3 s selected from a standard database. The sampling rate is 16 kHz and mixed signals are made synthetically based on an anechoic model. A ULA is used with a uniform spacing of 2 cm between adjacent sensors located at 1.5 m away from the sources. These results are averaged over 50,000 independent trials of each experiment. Computations are performed by a desktop PC with a Core 2 Duo CPU 2.33 GHz and 2 GB of RAM. The operating system is Windows 7 and all the codes are implemented in MATLAB R2008a.
5.1. Selection of b SFB sets

As stated in Section 5, incorporation of the bSFB set in the FVICA effectively reduces the running time while preserving the same separation quality compared to the ICA. To determine the bSFB set, we first choose a specific range of frequency bins from Figure 2 around the minimum point. Obviously, this range consists of more energetic frequency bins. We have shown that the given ranges in Table 1 are sufficient for finding the bSFB set for a different number of sources. The third column also shows the corresponding frequency bands in Hz according to the frequency resolution given in Eq. (16), which is a function of the FFT length \( R \) and the sampling rate \( f_s \).

\[
\Delta f = \frac{f_s}{R} = \frac{16000}{512} = 31.25 \text{ Hz}
\]  

(20)

Secondly, various bins in the ranges mentioned in Table 1 are considered as SFB sets to apply to the FVICA. Finally, by comparing the separation quality and running time of all sets, the bSFB set is found as presented in Table 2. It is worthwhile to mention that the running time of the proposed method consists of a) calculation time of transformation of mixed signals from the time to time-frequency domain using the STFT, b) separation of mixed signals in the IS stage, c) separation of mixed signals in the CS stage, and d) transformation of separated signals from the time-frequency to time domain using ISTFT.

Table 1. Range of frequency bins for finding the best SFB sets.

<table>
<thead>
<tr>
<th>( N )</th>
<th>Frequency bins range</th>
<th>Frequency bands (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8–30</td>
<td>125–937</td>
</tr>
<tr>
<td>3</td>
<td>5–30</td>
<td>125–937</td>
</tr>
<tr>
<td>4</td>
<td>5–30</td>
<td>125–937</td>
</tr>
</tbody>
</table>

To elaborate, the procedure of finding the bSFB set is explained in the following for \( N = 3 \).

A) According to Table 1, we consider different SFB sets for \( K = 2 \) in the range of 5 to 30 (125 to 937 Hz) and apply the FVICA to separate the signals of each set. Then the SFB set with the maximum separation quality is taken as the bSFB set for \( K = 2 \). This SFB is not the final bSFB set. In Figure 5, the amounts of SDR are shown for different SFB sets. As seen, the SFB set of \( \{16,17\} \) has presented the best quality.

![Figure 5](image-url)
B) We repeat step A in the range of 5 to 30 for all possible SFB sets with different numbers of components, i.e. $K = 3, \ldots, 26$, to find the bSFB sets whose results are shown in Figure 6.

C) The differences of SDR (dSDR), SIR (dSIR), and PESQ (dPESQ) between the FVICA for the bSFB sets for $K = 2$ to 26 and the ICA, and similarly the ratio of the running times (rRT) of the ICA to FVICA are shown in Figure 7, respectively.

From these results, we find the bSFB set for which the FVICA performs the same separation quality as the ICA. More specifically, considering the fact that the SDR and PESQ rather than the SIR play major roles in evaluating the separation quality, from Figures 7a, 7b, and 7c we select $K = 6$, for which dSDR (SDR of ICA minus SDR of FVICA) and dPESQ are very small (1.7 dB and 0.152, respectively). Then, from Figure 6, the bSFB set for $K = 6$ and $N = 3$ is $\{14, 15, 16, 17, 18, 19\}$, which is equivalent to 410 to 600 Hz.

One should note that separation qualities of both algorithms are so close that their difference is not in practice sensible by human hearing. More importantly, from Figure 7d the running time of the FVICA is 61.6 times less than that of the ICA, or, in other words, its separation speed is much faster than that of the ICA.

We similarly carried out steps A), B), and C) to find the bSFB sets and best frequency bands for $N = 2$ and 4, whose results are presented in Table 2.

<table>
<thead>
<tr>
<th>$N$</th>
<th>bSFB sets</th>
<th>Best frequency bands (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>[16, 17, 18, 19, 20]</td>
<td>470–630</td>
</tr>
<tr>
<td>3</td>
<td>[14, 15, 16, 17, 18, 19]</td>
<td>410–600</td>
</tr>
<tr>
<td>4</td>
<td>[11, 12, 13, 14, 15]</td>
<td>315–473</td>
</tr>
</tbody>
</table>

From a practical point of view, it is now sufficient for the reader to only apply the bSFB sets introduced in Table 2 to the FVICA for separating any given mixed speech signals with anechoic mixing model. This procedure is shown in Figure 8. It is important to note that the FVICA can also be applied to the mixed signals
(other than speech sources) of different applications whose energies are concentrated on the specific frequency bins. Clearly, for such cases the computations of the $b$SFB sets should be repeated according to Section 5.

5.2. Performance of FVICA for $b$SFB sets

In Figure 9, we compare the SDR versus running time of the FVICA with those of the ICA, FastICA, JADE, and SOBI. We choose $\eta = 0.08$ as the best step-size parameter for the ICA, which approximately maximizes the separation quality while minimizing the running time [29]. For example, according to Figure 9, the average running times of ICA and FVICA for $N = 2$ are 40.5 and 0.67 s, respectively. Thus, the rRT of ICA to FVICA is $40.5/0.67 = 60.45$, meaning that the FVICA separation time is about 60 times faster than that of the ICA. Similar results can be shown for $N = 3$ and 4. Accordingly, the average of dSDRs, dSIRs, dPESQs, and rRTs for $N = 2, 3, \text{and } 4$ are shown in Table 3. As the main goal of this work, the FVICA is almost 65 times faster than the conventional ICA while preserving the separation quality. The FVICA is also about 3 times faster than the FastICA with a 10 dB higher SDR and a similar improvement in separation quality and speed with respect to the JADE and SOBI.

![Figure 8. Flowchart of applying the FVICA to mixed speech signals.](image)

![Figure 9. Amounts of SDR versus the running time of FVICA, ICA, FastICA, JADE, and SOBI. Each graph shows 3 points for $N = 2, 3, \text{and } 4$, respectively.](image)

<table>
<thead>
<tr>
<th>Algorithms Criterion</th>
<th>SOBI</th>
<th>JADE</th>
<th>FastICA</th>
<th>ICA</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSDR (dB)</td>
<td>21.82</td>
<td>-13.25</td>
<td>-11.06</td>
<td>1.6</td>
</tr>
<tr>
<td>dSIR (dB)</td>
<td>-34.73</td>
<td>-15.2</td>
<td>-11.69</td>
<td>2.26</td>
</tr>
<tr>
<td>dPESQ</td>
<td>-1.6</td>
<td>-0.67</td>
<td>-0.5</td>
<td>0.12</td>
</tr>
<tr>
<td>rRT (times)</td>
<td>4.58</td>
<td>2.55</td>
<td>3.43</td>
<td>64.57</td>
</tr>
</tbody>
</table>

6. Conclusion
The new FICS method and corresponding FVICA algorithm was introduced in order to speed up the separation of anechoic mixed speech signals extremely without losing the quality. This was performed by 1) separating mixed signals iteratively in a SFB set and estimating mixing parameters, and 2) incorporating estimated

2053
parameters for signal separation in RFB sets in closed form. It was also shown that for speech signals separation can be performed in bSFB sets to avoid a large number of computations. Using simulation results, it was demonstrated that the FVICA is on average 65 times faster than the ICA, while the separation quality is preserved. The separation quality and speed of the FVICA is also better than those of the FastICA, JADE, and SOBI.

Acknowledgment
The authors would like to acknowledge financial support from the Iran Telecommunication Research Center (ITRC).

References


