METHODS

A permutation algorithm of frequency-domain blind source separation based on influence weights

Weihong Fu[*] and Tianqi Wu

School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi, China

[*]**Correspondence:**
Weihong Fu,
whfu@mail.xiidian.edu.cn

In the frequency domain, convolutive mixes with blind source separation can be successfully resolved. But the permutation issue in frequency-domain blind source separation needs to be resolved. We investigated the impact of frequency space and separation performance at each frequency bin on the amplitude correlation permutation algorithm with the goal of addressing the permutation ambiguity problem in frequency-domain blind source separation of convolutive mixtures, and we proposed an enhanced permutation algorithm. The improved algorithm uses spacing influence weight and performance influence weight to control the influence of the frequency bins sorted in the neighborhood on the frequency bins unsorted. Experiments have shown that the two influence weights are effective. Finally, blind source separation experiments are performed on the speech signals under the two convolutive mixing models and the simulated room mixing model. According to experiments, the increased signal to interference plus noise ratio of separated signals demonstrates that the improved algorithm outperforms the amplitude correlation permutation algorithm in terms of separation performance and robustness.

Keywords: convolutive mixture, frequency domain blind source separation, permutation, influence weight

Introduction

The process of separating many sources that have been blended together using just their observable mixtures and unidentified paths is known as "blind source separation"(1). As a rapidly developing emerging signal processing technology in the past 20 years, it has been widely used in digital communication, (2, 3) biomedical signal processing, (4, 5) speech signal processing, (6, 7) and other fields.

In recent years, the blind source separation of convolutive mixtures has become a research hotspot due to its great accordance with the current situation. The time-domain approach (8) and the frequency-domain method are its primary separating techniques. The Short Time Fourier Transform (STFT) converts convoluted mixed time-domain data into instantaneous mixed time-domain signals in the frequency domain. The instantaneous mixed signals in the frequency domain can be separated with the mature ICA algorithm, but the scaling and permutation ambiguities (9) must be resolved.

The separation matrix can be normalized to address the scaling ambiguity. The direction of arrival (DOA) technique (10) and the amplitude correlation algorithm between adjacent frequency bins (11) are the two primary methods for resolving permutation ambiguity (Murata algorithm). The former algorithm is more robust, but it has high environmental sensitivity and computational complexity. The latter algorithm exploits the amplitude dependence of adjacent frequency bins of the signals. It takes the sum of the amplitude correlations of adjacent frequency bins of the mixed signals as a cost function, maximizes it, and finds the corresponding permutation. Although this approach is computationally cheap, its performance is erratic and it lacks robustness. The current research shows that the Murata algorithm's poor robustness can be effectively improved by changing the impact factor of the frequency bins sorted in its neighborhood, thus improving its performance (12).

In this paper, we have studied the effect of frequency spacing and the separation performance at each frequency bin on the amplitude correlation permutation algorithm. The algorithm adjusts the influence of each frequency bin on the permutation result by spacing influence weight and performance influence weight. The validity of two influence weights is verified by experiments. An improved permutation algorithm is proposed with two influence weights considered simultaneously.

Frequency-domain blind source separation model of convolutive mixtures

Under the convolutive mixing model, the N independent sources $s_i(t)(i = 1, 2, \cdots, N)$ are transmitted, and the mixed signals received by the M sensors are $x_j(t)(j = 1, 2, \cdots, M)$. Its model can be expressed as:

$$x_j(t) = \sum_{i=1}^{N} \sum_{p=0}^{P-1} h_{ji}(p)s_i(t-p), j = 1, 2, \cdots, M \quad (1)$$

where $h_{ji}(p)$ denotes the impulse response of the source i and sensor j. When $p = 1$, the model is an instantaneous mixing model, and when $p > 1$, the model is a convolutive mixing model. P is the order of the impulse response function filter. Express the above expression as a vector form:

$$x(t) = \sum_{p=0}^{P-1} H(p)s(t-p) \quad (2)$$

where $s(t) = [s_1(t), s_2(t), \cdots, s_n(t)]^T$ and $x(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^T$ are source vectors and mixed signal vectors, respectively. $H(p)$ refers to the impulse response matrix with a delay of p, which can be expressed as:

$$H(p) = \begin{pmatrix} h_{11}(p) & h_{12}(p) & \cdots & h_{1n}(p) \\ h_{21}(p) & h_{22}(p) & \cdots & h_{2n}(p) \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1}(p) & h_{m2}(p) & \cdots & h_{mn}(p) \end{pmatrix} \quad (3)$$

A short-time Fourier transform (STFT) is performed on both sides of Equation (2) to obtain a frequency-domain mixing model of the convolved mixed signals:

$$X(f_k, \tau) = \tilde{H}(f_k)S(f_k, \tau) \quad (4)$$

where $S(f_k, \tau) = [S_1(f_k, \tau), S_2(f_k, \tau), \cdots, S_N(f_k, \tau)]^T$ and $X(f_k, \tau) = [X_1(f_k, \tau), X_2(f_k, \tau) \cdots, X_M(f_k, \tau)]^T$ are STFT signals of the sources and the mixed signals at frequency bin f_k, respectively, f_k represents the frequency bin index, and τ represents the time index. $\tilde{H}(f_k)$ represents the mixing matrix after the discrete Fourier transform, i.e.,

$\tilde{H}_{ij}(f_k) = DFT(h_{ij}), h_{ij} = [h_{ij}(0), h_{ij}(1), \cdots, h_{ij}(P-1)]$. The demixing model in the frequency domain obtained by Equation (4) is:

$$Y(f_k, \tau) = W(f_k)X(f_k, \tau) \quad (5)$$

where $W(f_k)$ is the demixing matrix at frequency bin f_k, and $Y(f_k, \tau) = [Y_1(f_k, \tau), Y_2(f_k, \tau), \cdots, Y_N(f_k, \tau)]^T$ is the N separated signals at frequency bin f_k, which can be restored to the time-domain signals by inverse short time Fourier transform (ISTFT).

Ambiguity problem

STFT is used in frequency-domain blind source separation to transform time-domain convolutive mixed signals into frequency-domain instantaneous mixed signals, and an instantaneous blind source separation technique is used to acquire the separated signals at each frequency bin. However, owing to scaling and permutation ambiguities of frequency-domain blind source separation, the mixing matrix $\tilde{H}(f_k)$ and the demixing matrix $W(f_k)$ in Equations (4) and (5) will satisfy:

$$W(f_k) = D(f_k)P(f_k)\tilde{H}^{-1}(f_k) \quad (6)$$

Where $D(f_k)$ is a random diagonal scaling matrix that captures scaling uncertainty, and $P(f_k)$ is a matrix of permutations that illustrates the permutation ambiguity. Different amplitude gains of the split signals at each frequency bin will result from scaling ambiguity. Permutation ambiguity can lead to sorting errors in the separated signals. Both will affect the algorithm's separation effect.

The separation matrix of each frequency bin can be normalized in order to remove scaling ambiguity (13):

$$W(f_k) \leftarrow diag[W^{-1}(f_k)]W(f_k) \quad (7)$$

Due to its low requirements for priori information and small computational complexity, the Murata algorithm is widely used for permutation ambiguity. However, the algorithm has poor robustness and unstable performance. The proposed algorithm effectively improves the disadvantage of the Murata algorithm by introducing the spacing influence weight and the performance influence weight, so that a better performance can be obtained.

Permutation algorithm

Correlation-based permutation algorithm

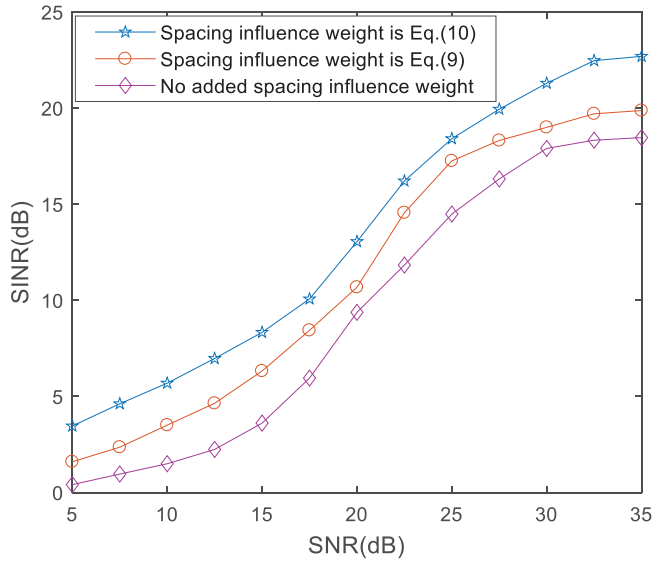In order to resolve the permutation ambiguity, the correlation-based permutation method (Murata algorithm)

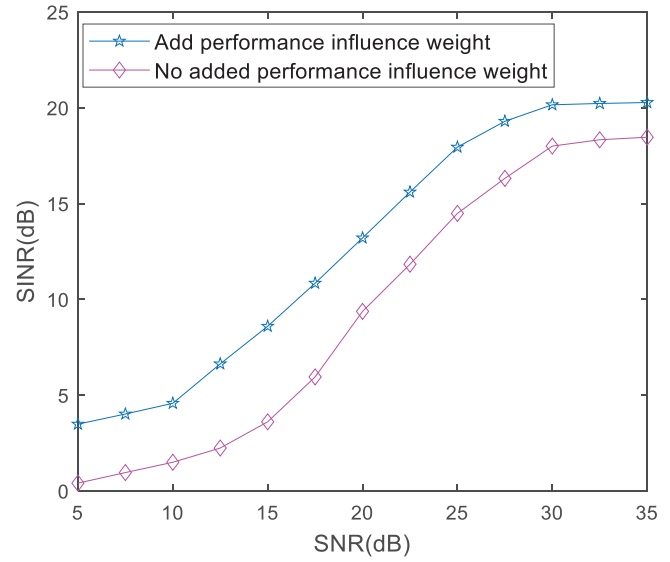**FIGURE 1 |** The impact of spacing influence weights on performance.

FIGURE 2 | The impact of spacing influence weights on performance.

sorts the separated signals by determining the amplitude correlation coefficient between each unsorted frequency bin and its surrounding frequency bins. When sorting the frequency bin f, we take the signals in a certain neighborhood $|g - f| \leq L$ as a reference and choose a permutation that can maximize the amplitude correlation coefficient. The frequency bin's most appropriate permutation is this one. This is how the algorithm works:

It is known that the permutation of the separated signals on the sorted frequency bins g is Π_g, and the permutation Π_f of the frequency bin f can be obtained by:

$$\Pi_f = \arg \max_{\Pi} \sum_{|g-f| \leq L} \sum_{i=1}^{N} cor\left(\left|Y_i^{\Pi}(f)\right|, \left|Y_i^{\Pi_g}(g)\right|\right) \quad (8)$$

where $Y_i^{\Pi}(f) = [Y_i^{\Pi}(f, 1), \cdots, Y_i^{\Pi}(f, \tau) \cdots Y_i^{\Pi}(f, T)]$ represents the separated signal i of the frequency bin f. Π is one of the permutations on frequency bin f. τ means time index and T represents time length. $Y_i^{\Pi_g}(g) = [Y_i^{\Pi_g}(g, 1), \cdots, Y_i^{\Pi_g}(g, \tau) \cdots, Y_i^{\Pi_g}(g, T)]$ indicates the separated signal i of the sorted frequency bins g, and $|\bullet|$ denotes gaining signal amplitude. From the above formula, the correct permutation at the frequency bin f can be obtained.

Improved permutation algorithm

A large number of experiments show that the Murata algorithm doesn't have a good performance, for it treats all the sorted frequencies in the neighborhood equally. Some frequency bins far away from the frequency to be sorted or with poor separation performance are also used as references. These low-reliability frequency bins will lead to permutation errors, which will make the whole algorithm less robust.

Aiming at the earlier discussed problems, we propose an improved algorithm.

If we only consider the effect of the adjacent frequency bin (neighborhood range $L = 1$), once the permutation of this frequency bin goes wrong, the permutation of all the frequencies subsequent will be incorrect. Therefore, the reference for permutation should be the sum of the correlation coefficients of the frequency bins within a certain neighborhood (neighborhood range $L \geq 2$). Also, the characteristic that the correlation among homologous signals is much higher than that among heterogeneous signals will gradually weaken as the neighborhood range increases. Thus, the spacing influence weight $\xi(g, f)$ should decrease as the neighborhood range increases. Therefore, $\xi(g, f)$ can be defined as (12):

$$\xi(g, f) = \frac{L - |g - f| + 1}{L} \quad (9)$$

The spacing influence weight proposed in (12) decreases linearly with the expansion of the neighborhood range. But the larger the frequency spacing $|g - f|$ is, the stronger the permutation interference between the frequency bins g and f is. Therefore, the effect of the frequency bins with small spacing should be highlighted as much as possible. Thus, we define the improved spacing influence weights as:

$$\xi(g, f) = \left(\frac{L - |g - f| + 1}{L}\right)^2 \quad (10)$$

In Equations (9) and (10), L is the neighborhood range, $g \in (f - 1, f - 2, \cdots, f - L)$ represents the frequency bins that have been sorted in the neighborhood, and f indicates the frequency bin to be sorted. The farther the frequency bin g is from the frequency bin f, the smaller the spacing influence weight $\xi(g, f)$ is.

TABLE 1 | Influence of different neighborhood ranges on algorithm performance.

Neighborhood range	2	3	4	5	6	7	8	9	10
SINR (dB)	20.24	22.49	23.41	24.12	24.67	23.76	23.81	23.59	23.64
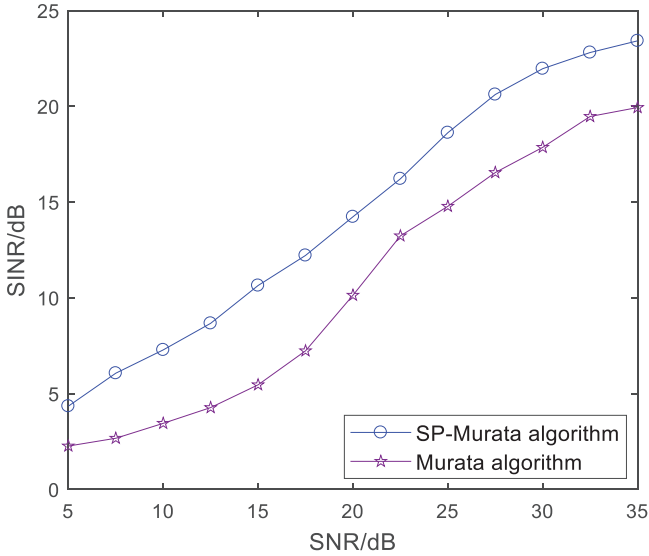
FIGURE 3 | Performance comparison of algorithms in 10-order mixed model.

The demixing matrix across frequencies has continuity, which can characterize the separation performance of each frequency bin (14). Generally speaking, for adjacent frequency points, if the signals are from the same source, they are more correlated, and the determinant of the demixing matrices at these frequency bins is continuous. When the determinant of the demixing matrix changes drastically, the correlation among signals from different sources exceeds that among signals from the same source. If frequency bins with poor separation performance are used as a basis for judging the permutation, it is bound to cause a decrease in the permutation effect.

Thus, the mean squared error corresponding to the frequency bin f is defined as follows:

$$\varphi(f) = \frac{1}{L} \sum_{|g-f| \leq L} \left(\left| \det(W(g)) \right| - \left| \det(W(f)) \right| \right)^2 \quad (11)$$

where $W(f)$ is the demixing matrix of frequency bin f, $\det(\bullet)$ represents the determinant operation of the matrix, $|\bullet|$ denotes the modulo operation. The larger the mean squared error $\varphi(f)$ is, the more intense the determinant changes, which means the separation performance of the frequency bin is poor. The smaller the mean squared error $\varphi(f)$ is, the better the continuity of the demixing matrix is, so the performance influence weight is inversely proportional to the mean squared error $\varphi(f)$. Therefore, we define the

performance influence weight at frequency bin f as:

$$\phi(f) = \frac{1}{\varphi(f)} \quad (12)$$

The improved decision formula based on the amplitude correlation permutation algorithm obtained by the above analysis is:

$$\Pi_f = \arg\max_{\Pi} \phi(f) \sum_{|g-f| \leq L} \xi(g,f) \sum_{i=1}^{N} cor\left(\left| Y_i^{\Pi}(f) \right|, \left| Y_i^{\Pi_g}(g) \right| \right) \quad (13)$$

Algorithm steps

Step 1: The time domain convolutive mixed signals should be converted into instantaneous mixed signals at each frequency bin using the STFT.

Step 2: The JADE algorithm is used to accomplish the instantaneous blind source separation at each frequency bin in order to produce the separated signals $Y(f, \tau)$ and the demixing matrix $W(f)$.

Step 3: Remove the signal's ambiguous scaling at each frequency bin in accordance with Equation (7).

Step 4: The separated signals at the first frequency bin f_1 are used as a reference for permutation, and the optimal neighborhood range L_{max} is given.

Step 5: Starting from the second frequency bin f, if $|f - f_1| < L_{max}$, the neighborhood range $L = |f - f_1|$, otherwise the neighborhood range $L = L_{max}$.

Step 6: Calculate $\xi(g,f)$ and $\phi(f)$ according to Equations (10) and (12), and solve the permutation problem of frequency bin f according to Equation (13).

Step 7: The frequency-domain signals are restored to the time-domain signals by ISTFT, and the estimation of the source signals is completed.

Simulation experiments and analysis

In order to verify the performance of the proposed ambiguity, the algorithms before and after the improvement are compared by simulation experiments. Two sources, two sensors, and an impulse response function filter with an order of 10 are present. The source signals for the simulation experiment were two voice signals taken from the Hiroshi Sawada site of the NTT Communication Science Laboratory.
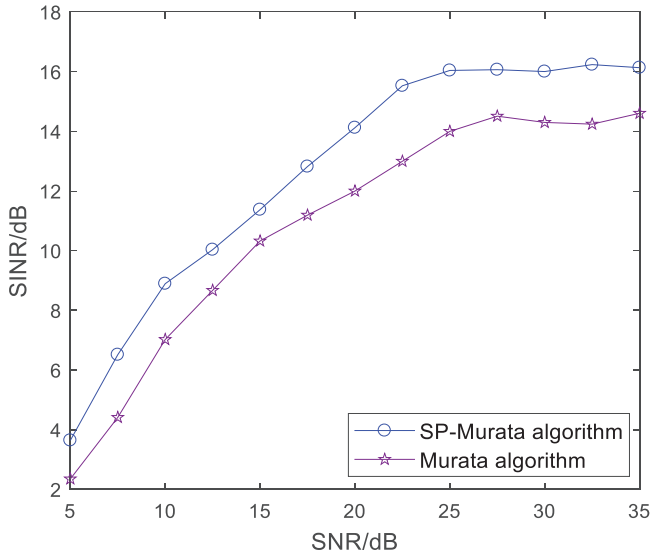
FIGURE 4 | Performance comparison of algorithms in 25-order mixed model.
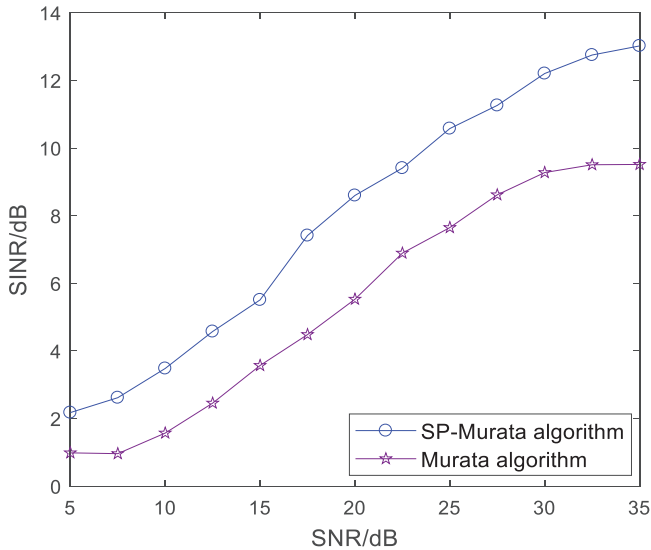

FIGURE 5 | Performance comparison of algorithms under the simulated room mixing model.

The signal to interference plus noise ratio (SINR) is used to estimate the separation effect, defined as:

$$SINR = 10 \lg \frac{\|s_{target}(n)\|^2}{\|y(n) - s_{target}(n)\|^2} \quad (14)$$

where $s_{target}(n)$ represents the target signal, and $y(n)$ denotes the separated signal.

The improved algorithm in this paper is called SP-Murata. We test its performance under different load weights. The SINR is applied in the examination to measure the separation effect of the improved algorithm.

When the performance influence weight $\varphi(f) = 1$ and the neighborhood range $L = 5$ are known, we test the two spacing influence weights $\xi(g,f)$ separately. As shown

in **Figure 1**, compared with the Murata algorithm, both the spacing influence weight given in (12) [Equation (9)] and that proposed in this paper [Equation (10)] can improve the separation performance to a certain degree. The spacing influence weight proposed in this paper has the best performance among the three algorithms. Compared with the Murata algorithm, the proposed algorithm can improve the SINR by 2−3 dB, which is 1−2 dB better than the algorithm in (12). Therefore, in the subsequent experiments, $\xi(g,f) = \left(\frac{L-|g-f|+1}{L}\right)^2$ is selected for the spacing influence weight.

When the spacing influence weight $\xi(g,f)=1$ and the neighborhood range $L = 5$, we test whether the addition of the performance influence weight $\varphi(f)$ could improve the separation performance. **Figure 2** shows that compared to the Murata algorithm, the performance influence weight $\varphi(f)$ can somewhat enhance the performance of the algorithm. This can prove the effectiveness of performance influence weight $\varphi(f)$ on algorithm improvement. Because both types of influence weights can improve the performance of the algorithm, we simultaneously use them to improve the Matura algorithm. What needs to be determined is the optimal neighborhood range in permutation, so the following experiment is performed:

It can be seen from the above analysis that the algorithm's performance has improved under the action of $\xi(g,f)$ and, $\varphi(f)$, respectively. Therefore, the two influence weights are simultaneously applied to the Murata algorithm. In order to find the optimal neighborhood range of the improved algorithm, we compare the algorithm's performance in different neighborhood ranges without noise. The results are shown in **Table 1**.

It shows that the SINR of the algorithm increases with the increase in the neighborhood range. When the neighborhood range reaches a certain value $L = 6$, the performance of the algorithm tends to be stable. When the neighborhood range is 6, the two advantages of the algorithm performance and the minimum calculation amount can be considered at the same time. Therefore, the value of the optimal neighborhood range is set to $L = 6$ in the following simulated trials.

Then, using the 10-order filter convolutive mixing model, the 25-order filter convolutive mixing model, and the simulated room mixing model, we evaluate the performance of the suggested algorithm under these three simulation scenarios (15). In the simulated room mixing model, we construct a simulated room of $5m \times 5m \times 2.5m$, and the impulse response function filter's order can be 400 or higher. To test the mixed signal, different SNRs of Gaussian white noise are applied. In **Figures 3–5**, the average SINR of the separated signals is displayed.

Figures 3–5 show the performance comparison between the suggested approach and the Matura algorithm for the three different convolutive mixing models: the 10-order model, the 25-order model, and the simulated

room model. The figures show that, in comparison to the Murata algorithm, the SP-Murata algorithm performs much better under these three convolutive mixing models. The SINR reflects this. The performance of the algorithm can be increased by $2-3$ dB for both the 10-order and 25-order convolutive mixing models. The performance of the algorithm can be improved by $3-4$ dB in the simulated room model.

Conclusion

The frequency domain solution of the permutation ambiguity problem of multiple frequency bins is necessary for the blind source separation technique of convolutive mixes. This study presents a frequency-domain blind source separation permutation method with improved amplitude correlation based on effect weights. The algorithm considers the effect of adjacent frequency bin space, separation performance of each frequency bin, and neighborhood ranges, and solves the poor robustness of the Matura algorithm. In this paper, we verify the improvement of the proposed two types of influence weights on the performance and finally obtain the SP-Matura algorithm. On the basis of this, various convolutive mixing models are used to test the algorithm's performance. The outcomes demonstrate that the suggested algorithm can substantially enhance separation performance.

Author contributions

WF contributed significantly to analysis and manuscript preparation. TW performed the experiment. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zhao L, Beibei Z, Wu XP, Zhang C, Zhou B. A permutation algorithm based on dynamic time warping in speech frequency-domain blind source separation. *Speech Commun.* (2017) 92: 132–41.

2. Chen AJ, Yang QS, Wang J, et al. Blind source separation of LFM signals using time frequency domain sparseness. *Telecommun Eng.* (2015) 55:61–7.

3. Haile MA, Dykas B. Blind source separation for vibration-based diagnostics of rotorcraft bearings. *J Vibrat Control.* (2016) 22: 3807–20.

4. Negro F, Muceli S, Castronovo AM, Holobar A, Farina D. Multichannel intramuscular and surface EMG decomposition by convolutive blind source separation. *J Neural Eng.* (2016) 13: 26–7.

5. Ali A, Benjamin R, Justin R. Blind deconvolution using convex programming. *IEEE Trans Inform Theory.* (2014) 60: 1711–32.

6. Tariqullah J, Haseeb Z, Ruhulamin K, et al. A blind source separation approach based on IVA for convolutive speech mixtures. *Proceedings of the 8th Computer Science and Electronic Engineering Conference (CEEC), Colchester, United Kingdom, Sep. 2016.* Colchester: (2016). p. 140–5.

7. Chabriel G, Kleinsteuber M, Moreau E, Shen H, Tichavsky P, Yeredor A, et al. Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications. *IEEE Signal Process Mag.* (2014) 31:34–43.

8. Jia ZC, Han DW, Chen L, et al. Convolutive blind separation algorithm based on complex Givens matrix and bat optimization. *J Commun.* (2016) 37:107–17.

9. Saito S, Oishi K, Furukawa T. Convolutive blind source separation using an iterative least-squares algorithm for non-orthogonal approximate joint diagonalization. *IEEE/ACM Trans Audio Speech Lang Process.* (2015) 23:2434–48.

10. Saruwatari H, Saruwatari H, Kawamura T, Nishikawa T, Lee A, Shikano K. Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. *IEEE Trans Audio Speech Lang Process.* (2006) 14:666–78.

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

12. Bo XL, He YG, Yin B, Fang G, Fan X, Li Z, et al. Algorithm to eliminate permutation of frequency domain blind source separation based on influence factor. *Aata Electron Sin.* (2014) 42: 360–5.

13. Wang Z, Bi G. A time-frequency preprocessing method for blind source separation of speech signal with temporal structure. *Proceedings of the 10th International Conference On Information, Communications and Signal Processing.* New York, NY: (2015). p. 1–6.

14. Nesta F, Svaizer P, Omologo M. Convolutive BSS of short mixture by ICA recursively regularized across frequencies. *IEEE Trans Audio Speech Lang Process.* (2011) 19:624–37.

15. Pham DT, Serviere C, Boumaraf H. Blind separation of speech mixtures based on nonstationarity. *Proceedings of the International Symposium on Signal Processing and ITS Applications, Paris, France, July 2003.* (Vol. 2), Paris (2003). p. 73–6.