Journal of Computational Information Systems 9: 18 (2013) 7369–7379
Available at http://www.Jofcis.com

Performance Evaluation for Frequency Domain Blind Source Separation Algorithms [⋆]

Yueyue NA [1,*], Bianfang CHAI [2]

[1] *Department of Computer Science, Beijing Jiaotong University, Beijing 100044, China*

[2] *Department of information engineering, Shijiazhuang University of Economics, Shijiazhuang 050031, China*

Abstract

Blind source separation (BSS) aims at extracting original source signals from their mixed observations, it has many potential applications in different disciplines. In order to develop better BSS algorithms, how to evaluate the algorithm performance becomes a problem worthy to be investigated. In this paper, we mainly focus on the evaluation problem for frequency domain BSS algorithms. First, the uniform energy flow network is calculated from the mixing and the demixing system, or estimated from the original and the estimated source signals in frequency domain, then, signals are decomposed according to their energy flow, and several performance indices are derived from the decomposition. The proposed method is especially suitable for BSS performance evaluation in simulated environments, experimental results show that the proposed method is more accurate but less computational expensive than the state-of-the-art evaluation algorithms.

Keywords: Blind Source Separation; Performance Evaluation; System Identification

1 Introduction

The problem of extracting individual source signals from their mixed observations has been studied for many years, which is often known as the blind source separation (BSS) problem. The word "blind" refers to the fact that neither the source prior information, nor the mixing environment is known by the user in advance. There are many potential applications for BSS techniques, such as speech enhancement, robust speech recognition [1], separating superimposed moving images [2], high speed train noise component separation [3], etc.

Many BSS algorithms based on different assumptions have been proposed for different applications. For example, Independent component analysis (ICA) [4] is often used to solve the instantaneous mixing problem, the only assumption made by ICA is that source signals are mutually independent. Frequency domain blind source separation (FDBSS) algorithms [1, 5], and

[⋆] Project supported by the Hebei province Science and Technology Planed Projects (Grant No. 11213564, and Grant No. 13210702D).

[*] Corresponding author.
Email address: 08112069@bjtu.edu.cn (Yueyue NA).

1553–9105 / Copyright © 2013 Binary Information Press
DOI: 10.12733/jcisP0159
September 15, 2013

independent vector analysis (IVA) algorithms [3, 6, 7] are often used to solve the speech and audio separation problem. Since speech and audio signals are convolutively mixed in normal room environment, performing separation in frequency domain can transfer time domain convolution to frequency domain dot product, so that the problem can be simplified. In the case that the number of sensors is less than the number of sources, i.e. the underdetermined problem, the time-frequency masking approach [8] can be used on the assumption that source signals are not only independent but also sparse.

As more and more BSS algorithms are developed, how to evaluate their separation performance becomes a problem. Usually, the separated signals suffer from permutation, scaling, time delay, convolution indeterminacies, its not easy to evaluate the separation performance by comparing the waveform between the original and the separated signals directly, instead, the signal-to-interference ratio (SIR) [1] is often used for the evaluation purpose. Supposing s_{target} is the true source signal, and e_{interf} is the interference from other sources, then, SIR can be calculated as (1), where $\|s(t)\|^2 = \sum_t s^2(t)$ calculates the energy of a sequence. The higher SIR value means more target component and/or less interference component exists in the estimated signal, which means the higher separation performance.

$$SIR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf}\|^2} \tag{1}$$

SIR is probably the most widely used performance index for BSS evaluation, which can be calculated both in time domain [1, 9] and in frequency domain [10]. In [11], the estimated signal was decomposed into different components by subspace projection, and the performance indices were calculated from the decomposed components. The advantages of this approach are: First, the separation performance can be evaluated only from the source and the estimated signals, no mixing and demixing models are assumed in the evaluating procedure, so, it can be used for a variety of BSS algorithms; Second, three other performance indices: signal-to-distortion ratio (SDR), signal-to-noise ratio (SNR), signal-to-artifact ratio (SAR) can be derived from the decomposed components. Recently, the method in [11] was further updated by incorporating humans subjective assessment [12].

Although the subspace projection approach [11] is broadly used, it still has a drawback: for speech and audio separation tasks, filters used to model the mixing and demixing environments are usually thousands of taps long, so, thousands of base vectors, which are generated from the original source and the noise signals with different time delays, are needed to span the target, the interference, and the noise subspaces, such high complexity makes the subspace projection difficult to be calculated. The Matlab implementation [13] of the subspace projection approach solves this problem in two ways: first, fewer base vectors are generated to span the approximated subspaces, then, FFT is used to calculate the projections in frequency domain conveniently. However, approximated subspaces may bring extra artifacts to the evaluation, and very long FFT is required when the input and output signals are long.

In this paper, we focus on the performance evaluation problem for frequency domain BSS algorithms, and two approaches are presented: first, in the simulated mixing experimental environment, the original source signals and the mixing filters are usually known in advance, so, the two pieces of information can be utilized by the evaluation procedure. Inspired by the work in [11], we perform signal decomposition and performance evaluation in frequency domain; Second, when only the original sources and the estimated signals are available, a system identification method is proposed to estimate the system frequency responses, then signals can be decomposed based on the estimation. The proposed methods are easy to use, and the time complexity is low, all these approaches are integrated in a uniform platform for BSS research and real-world application purpose, which is available for public [14]. The rest of this paper is organized as follows: the way to perform signal decomposition and performance evaluation in frequency domain is depicted in section 2. The method for system frequency response estimation is introduced in section 3, which can be used when the mixing environment is unknown. Then, experimental results and comparisons are reported in section 4. At last, we conclude this paper in section 5.

2 Performance Evaluation in Frequency Domain

2.1 Performance indices

In [11], the estimated source signal y was first decomposed into different components as depicted in (2), the decomposition was made by projecting y into different subspaces spanned by the original sources and the noise signals.

$$y = s_{target} + e_{interf} + e_{noise} + e_{artif} \tag{2}$$

In equation (2), s_{target} is the component of the true source signal in y, e_{interf} is the interference from other sources, e_{noise} is the component of sensor noise, and e_{artif} is the artifact resulted from the separation and evaluation algorithms. In addition to SIR in (1), three other performance indices were defined upon the decomposition in (2):

$$SDR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf} + e_{noise} + e_{artif}\|^2} \tag{3}$$

$$SNR = 10\log_{10}\frac{\|s_{target} + e_{interf}\|^2}{\|e_{noise}\|^2} \tag{4}$$

$$SAR = 10\log_{10}\frac{\|s_{target} + e_{interf} + e_{noise}\|^2}{\|e_{artif}\|^2} \tag{5}$$

In human's subjective perspective, signal-to-distortion ratio (SDR) can be considered as the overall assessment to the separation algorithm; Signal-to-interference ratio (SIR) evaluates the interference rejection ability of the algorithm, which is the most important performance index in BSS evaluation; Signal-to-noise ratio (SNR) evaluates the influence from the sensor noise, in this paper we simply set $e_{noise} = 0$, and $SNR = \infty$, just the same as the Matlab implementation [13] of the subspace projection method did; Signal-to-artifact ratio (SAR) is used to measure the amount of artifacts introduced by the separation algorithm, which is usually shown as the burbling effect and the musical noise [11].

2.2 Signal decomposition in frequency domain

Once the decomposition in (2) is calculated, performance indices can be easily derived from the decomposed components according to (3), (1), (4), and (5), so, different decomposition methods

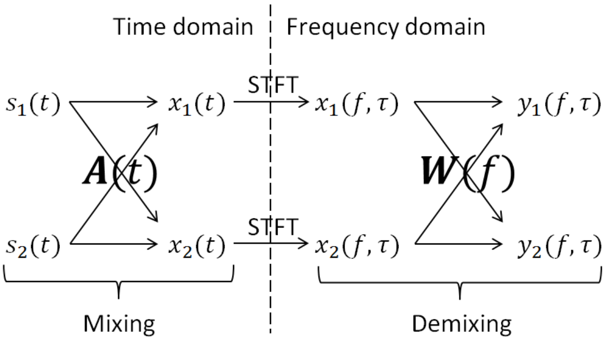

Fig. 1: FDBSS mixing and demixing system

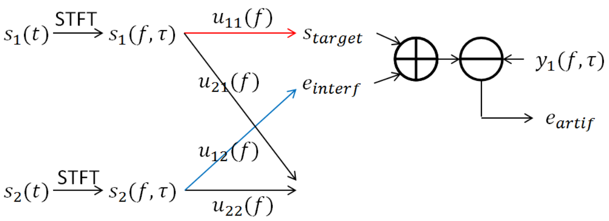

Fig. 2: Signal decomposition in frequency domain

other than subspace projection can also be used. In frequency domain BSS, the mixing filters $\boldsymbol{A}(t)$ is usually known in simulated experiments, and the sensor signals are artificially generated by convolving the source signals with the mixing filters. An example of 2×2 mixing and demixing system is shown in Fig. 1: the original source signals $\boldsymbol{s}(t) = [s_1(t), s_2(t)]^T$ is convolutively mixed by the mixing filters $\boldsymbol{A}(t)$, then the FDBSS algorithm transforms the sensor signals $\boldsymbol{x}(t)$ into time-frequency domain via short time Fourier transform (STFT), here index t represents time domain data, and index f represents frequency domain data. After the demixing system $\boldsymbol{W}(f)$ is calculated, the estimated sources can be derived as: $\boldsymbol{y}(f,\tau) = \boldsymbol{W}(f)\boldsymbol{x}(f,\tau)$, where f is the frequency bin index, τ is the STFT frame index.

When we want to evaluate the separation performance, the convolution theorem can be applied, and the mixing, demixing system in Fig. 1 can be combined into a uniform system in frequency domain as: $\boldsymbol{U}(f) = \boldsymbol{W}(f)\boldsymbol{A}(f)$, which is depicted in Fig. 2. Taking s_1 as example, supposing its corresponding output is y_1, then, the signal can be decomposed as (6)-(8), according to the energy flow network in Fig. 2. Please notice that, because of the permutation indeterminacy, the corresponding output of s_1 may flow to other channels other than y_1, so, in real applications we try all possibilities and select the best result. Signals in other input and output channels can also be decomposed similar to this example. According to the Parsevals theorem, signal energy preserves in frequency domain, so, energy can be directly accumulated upon the time-frequency data. After the decomposition, the performance indices can be calculated by (3), (1), (4), and (5). The difference between our approach and the subspace projection approach is that the mixing system is needed in the proposed method, however, this requirement can be definitely fulfilled in simulated experiments. Since there is no need to perform orthogonal projections in very high dimensional subspaces, the time complexity of the proposed method is very low.

$$s_{target} = u_{11}(f)s_1(f,\tau) \tag{6}$$

$$e_{interf} = \sum_{n \neq 1} u_{1n}(f) s_n(f, \tau) \tag{7}$$

$$e_{artif} = y_1(f, \tau) - \sum_{n} u_{1n}(f) s_n(f, \tau) = y_1(f, \tau) - (s_{target} + e_{interf}) \tag{8}$$

2.3 Input SIR

The signal-to-interference ratio improvement (SIRI) is usually used to measure the amount of quality improvement obtained by the BSS algorithm, which can be calculated according to (9) [9, 15]. In (9), output SIR means the result in (1), and input SIR measures the signal-to-interference ratio of the sensor signals. In the literature, input SIR of source n is usually calculated according to (10) [15], or other similar policies [9]. These approaches assume that the target component of s_n can be approximated by convolving s_n with the mixing filter a_{nn} or a_{Jn}, where J is the reference sensor index [9]. Because of the permutation indeterminacy, the problem of (10) is that the target mixing filter corresponding to s_n is unknown, simply using the diagonal filter a_{nn}, or designating a reference sensor a_{Jn} may yield wrong results.

$$SIRI = OutputSIR - InputSIR \tag{9}$$

$$InputSIR_n = 10 \log_{10} \frac{\|a_{nn} * s_n\|^2}{\sum\limits_{n' \neq n} \|a_{nn'} * s_{n'}\|^2} \tag{10}$$

To solve this problem, we first define the energy flow matrix $\boldsymbol{P}$, whose entries are calculated as $p_{nm} = \|a_{mn}(f)s_n(f, \tau)\|^2$ represent the energy flow from source n to sensor m. In $\boldsymbol{P}$, the column sum $p_m = \sum_n p_{nm}$ is the total energy received by sensor m. Then, the SIR flow matrix $\boldsymbol{Q}$ can be calculated as: $q_{nm} = p_{nm}/(p_m - p_{nm})$, which assuming p_{nm} is the target energy, and the rest part of energy at sensor m is the interference. Since the real target energy is unknown, the best result can be selected as the input SIR of source n: $InputSIR_n = 10 \log_{10} \max_m q_{nm}$. This approach can also be used to calculate output SIR, the only difference is that the transfer system is now $\boldsymbol{U}(f)$, instead of the mixing system $\boldsymbol{A}(f)$. Since e_{artif} is not decomposed, the computational complexity of this approach is slightly lower than the method in the section 2.2.

3 Evaluation by System Identification

In the situation of [11], the only available data are the original sources $\boldsymbol{s}$ and the output signals $\boldsymbol{y}$, the uniform system $\boldsymbol{U}(t)$ cannot be derived as the mixing system $\boldsymbol{A}(t)$ is unknown. In order to decompose signals according to the method in section 2, $\boldsymbol{U}(t)$ must be properly estimated first, and the method of system identification can be used for this purpose.

In Fig. 3, supposing s is the input of an unknown linear time invariant (LTI) system, and y is the corresponding output, usually, the output is disturbed by unknown system noise e, which is independent from s. This procedure can be formulated in (11) in time domain, and (12) in frequency domain, where $h(t)$ and $h(f)$ are the impulse response and corresponding frequency

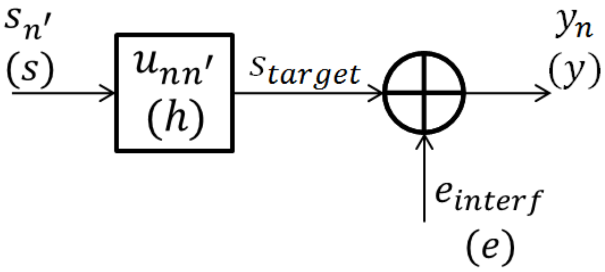

Fig. 3: System identification

response of the system, respectively. The task of system identification is estimating the system frequency response $h(f)$ in the noisy environment of Fig. 3 from the system input and output only.

$$y(t) = h(t) * s(t) + e(t) \tag{11}$$

$$y(f, \tau) = h(f)s(f, \tau) + e(f, \tau) \tag{12}$$

When $h(t)$ is finite duration impulse response (FIR), the system frequency response can be estimated according to (13) by exciting the system with a white noise signal s [16, 17].

$$h(f) = \frac{\sum\limits_{\tau} y(f, \tau)s^*(f, \tau)}{\sum\limits_{\tau} |s(f, \tau)|^2} \tag{13}$$

Comparing the model in Fig. 3 with the FDBSS uniform system in Fig. 2, we can find that any source and estimated signal pair $s_{n'}$ and y_n meets this model. In FDBSS simulated experiments, the source signal $s_{n'}$ is known in advance, and the estimated signal y_n is outputted by the BSS algorithm, the frequency response of the LTI system is $u_{nn'}$, which is corresponding to the filter $u_{nn'}(f) = \sum_m w_{nm}(f)a_{mn'}(f)$ in Fig. 2. The system is interfered by the additive noise e_{interf}, which comes from other sources, obviously, $s_{n'}$ and e_{interf} are mutually independent.

To estimate the system frequency response, the system must be excited by white noise, this is due to white noise can be considered as broad band signal, and its spectrum is "white", i.e. energy components exist in the full frequency band, so, frequency response in all frequency bins can be properly estimated. For BSS algorithms, the original source signal is usually speech or audio, which may yield zero energy in some frequency bins, however, this will not affect its application in our problem, since the signal needs to be decomposed is the same as the one used to estimate the frequency response. To make sure the frequency response can be estimated properly, we slightly modify the method in (13) to (14), the constraint condition $\sum_\tau |s_{n'}(f, \tau)|^2 \geq 1$ is used to keep the estimated frequency response stable, it means that the numerator should not be magnified by the denominator in (14). After the uniform system $\boldsymbol{U}$ is estimated, signals can be decomposed according to the method in section 2.

$$u_{nn'}(f) = \begin{cases} (\sum\limits_{\tau} y_n(f, \tau)s^*_{n'}(f, \tau))/(\sum\limits_{\tau} |s_{n'}(f, \tau)|^2) & \sum\limits_{\tau} |s_{n'}(f, \tau)|^2 \geq 1 \\ 0 & otherwise \end{cases} \tag{14}$$

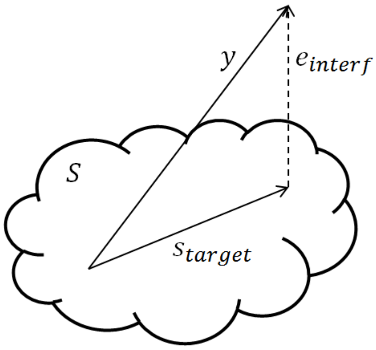

Fig. 4: Least square method and subspace projection

As pointed out in [16], the system identification approach in (13) and (14) can be considered as a least square method in frequency domain, which minimizing $\|y(t) - s(t) * u(t)\|^2$ by selecting proper $u(t)$. Supposing $S = span\{s(t), s(t-1), ..., s(t-L)\}$ is the subspace spanned by the delayed versions of the source signal s, by noticing that $s(t)*u(t) = s_{target} \in S$ and $e_{artif} = y(t)-s(t)*u(t)$, the least square method can be visualized as Fig. 4. On the other hand, the subspace projection method in [11] can be considered as: looking for a vector s_{target} in the subspace S, having the distance between s_{target} and y minimized, just as the same meaning as Fig. 4. From these two explanations we can see that the subspace projection approach [11] and the system identification approach [16, 17] are equivalent in theory.

4 Experiments

All experiments were carried out on a uniform platform for BSS research and application. This platform is developed in Java, some frequently used source separation algorithms have already integrated in the system, and the proposed methods in this paper are also implemented for evaluation purpose. The source code is available for public, please visit [14] for more information.

4.1 System identification algorithm evaluation

First of all, to verify the reliability of the frequency response estimation method in (14), a simulated environment in Fig. 3 was established. The system was excited by an artificially generated Gaussian white noise s, and the interference e was set to another Gaussian white noise, which was independent form the excitation. The mean square error $err = \sum_{l=1}^{L}(h(l) - \hat{h}(l))^2/L$ between the real impulse response h and the estimated impulse response $\hat{h}$ was used to evaluate the algorithm performance, where L is the impulse response length. The experimental results is shown in Fig. 5, we can see that the mean square error decreases as the number of STFT frames increases, moreover, this algorithm is very robust, for the application in BSS performance evaluation, acceptable mean square error can be achieved even in very low SIR environment.

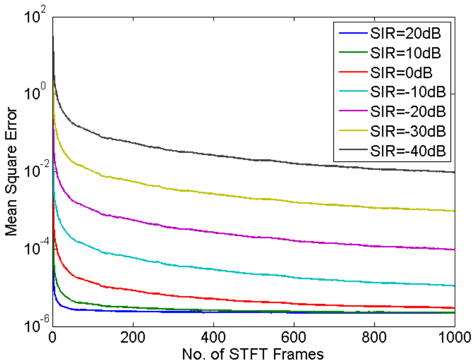

Fig. 5: System identification method evaluation

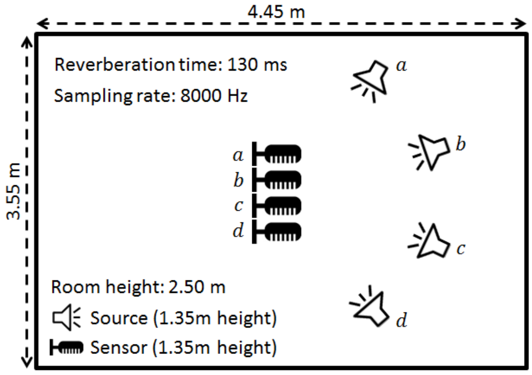

Fig. 6: The simulated environment

4.2 Comparison of BSS evaluation results

The second experiment compares the results outputted by different evaluation approaches. The simulated mixing environment was established according to Fig. 6 [18], mixing filters were generated by the image method [19], and the dataset from [18] was used as source signals. Two source separation algorithms were tested, including the subband subspace IVA algorithm (SSIVA) [3], and the complex valued ICA [4] plus sequentially align the signal envelop to correct the permutation ambiguity (ICA+SA). In all separations, STFT frame size was set to 1024, STFT frame overlap was set to 7/8, and FFT block size was set to 2048. Three evaluation methods were compared, including the bss_eval toolbox [13], which is the implementation of the subspace projection method in [11], with allowed time delays were set to 512 and 2048 (SP), the frequency domain decomposition method in section 2 (FDD), and the system identification method in section 3 (SI). The average perform indices are shown in Table 1.

In Table 1, the SIR values of all evaluation algorithms are less than 2dB of difference, this means that fairly assessments can be made by these evaluation algorithms. For frequency domain BSS algorithms, the separation performance will greatly affected by the permutation ambiguity [9, 15, 20]. The SIR values of the SSIVA approach are relatively high, it means that the permutation problem can be well solved by this approach. However, low SIR of the ICA+SA approach in the three and four sources cases means that separations were failed in these configurations, this is due to the permutation error in one frequency bin will propagate to other frequency bins in the sequentially align depermutation approach.

The SAR values between the SP and the SI approaches are close, since these two approaches are equivalent in theory. However, SAR values given by the FDD method are much higher. From (5) we can see that the energy of algorithm artifact heavily affects the value of SAR, the e_{artif} given by the FDD approach has higher accuracy since it is decomposed strictly according to the mixing-demixing energy flow network in Fig. 2. After outputted as wave file, we can see that the decomposed e_{artif} is made up of two parts: the tiny reverberation effect from the original source, and the STFT windowing effect at the beginning and the end of the signal. When comparing the simulated mixing network in Fig. 1 and the evaluating network in Fig. 2, it is easy to find that FDDs artifact is resulted from the difference between the time domain convolution and the frequency domain dot product of the mixing system in Fig. 1 and Fig. 2. In our experiment, if the mixing procedure in Fig. 1 is carried out in frequency domain, $e_{artif} \approx 0$ and $SAR \approx \infty$ will be outputted by the FDD approach. On the other hand, the artifact of the SP approach may result from the insufficiently allowed time delays (this can be seen that the SAR of 2048 taps

Table 1: BSS performance evaluation results

		2 sources (a, d)			**3 sources (a, b, d)**			**4 sources (a, b, c, d)**		
Separation algorithm	**Evaluation algorithm**	**SDR**	**SIR**	**SAR**	**SDR**	**SIR**	**SAR**	**SDR**	**SIR**	**SAR**
SSIVA	**SP(512)**	13.06	20.76	13.92	9.64	16.27	10.90	9.17	16.95	10.11
	SP(2048)	13.42	20.16	14.51	9.96	15.25	11.68	9.42	15.08	10.99
	FDD	19.24	19.53	31.37	15.67	15.80	31.20	16.37	16.49	32.41
	SI	13.32	20.63	14.31	10.58	16.54	11.94	9.42	15.08	10.99
ICA+SA	**SP(512)**	12.88	18.88	14.75	1.40	1.95	10.62	3.90	7.96	10.31
	SP(2048)	13.17	18.43	15.29	1.36	2.17	11.35	4.22	7.38	11.15
	FDD	17.64	17.90	31.06	1.70	1.71	31.43	7.86	7.90	34.21
	SI	13.17	19.02	14.90	0.99	1.60	12.40	4.98	7.61	12.06

time delay is higher than 512 taps in Table 1), and the artifact of the SI approach may result from the frequency response estimation error. In addition, when source signals are not strictly independent with each other, the SP and the SI approaches may yield improper decompositions. For an exaggerated example, supposing all source signals are identical, correct decompositions still can be made according to the energy flow network in Fig. 2 if the demixing system still can be estimated by the separation algorithm, however, the other two approaches will not work in this extreme circumstance.

4.3 Computational complexity analysis

To analyze the computational complexity, we first declare some required notations: let N, M, T, L represent source number, sensor number, STFT frame number, mixing filter length, respectively. In the subspace projection approach (SP), the key operation is calculating the projection of an estimated signal to the subspace spanned by the delayed versions of source signals, in [13], the projection is done by the least square method. In the least square approach, the inversion of the inner product matrix of size $N(L-1) \times N(L-1)$ must be calculated, whose complexity is $O(N^3L^3)$. At last, performance indices between each pair of signals should be calculated, so, the final time complexity is $O(N^5L^3)$. To calculate the SIR improvement (SIRI) introduced in section 2.3, input SIR and output SIR must be calculated first, which have the complexity of $O(MNT)$ and $O(N^2T)$, respectively. So, for the (over)determined problem ($M \geq N$), the final time complexity is $O(MNT)$. In the frequency domain decomposition approach (FDD) introduced in section 2.2, only the frequency domain dot product and addition operations are required, the time complexity is $O(N^2T)$. In the system identification approach introduced in section 3, frequency responses of the uniform system must be estimated first, and the complexity of (14) is also $O(N^2T)$, so, the total time complexity of this approach is $O(N^2T)$.

To compare the time complexity of different performance evaluation algorithms, dataset from [21] was used, signal sampling rate was 8000 Hz, wave file length was 6.25 seconds, and all experiments were carried out on a PC with 2.2GHz CPU and 2GB memory. Experimental results are shown in Fig. 7. From Fig. 7 we can see that when the number of sources is not too large, the evaluation methods proposed in this paper have nearly linear time complexity, the SIRI approach has lowest complexity, since only the input and the output SIR are calculated. The SI approach has higher complexity than the FDD approach, since it needs to estimate system frequency responses before performing signal decomposition. The SP approach has nearly exponential time complexity, however, when the number of sources is not too large, the SP approach still can perform evaluation in an acceptable time cost.

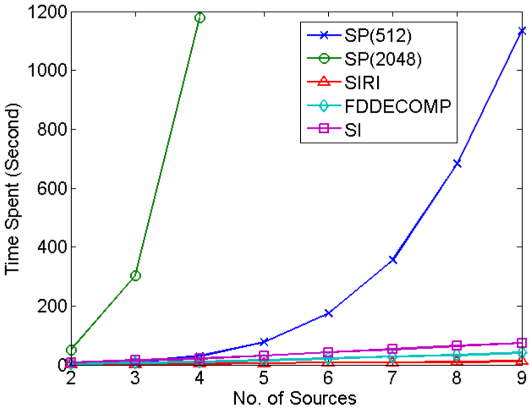

Fig. 7: Time complexity comparison

5 Conclusions

In order to develop better source separation algorithms, performance assessment indices are required to evaluate the BSS algorithm performance. In this paper, we mainly focus on the evaluation problem for frequency domain BSS algorithms, and a frequency domain signal decomposition policy is proposed based on the work in [11]. Compared with the existing techniques, the proposed methods are easy to use, and have lower time complexity. Since the mixing environment is required, the proposed decomposition method is mainly used in the simulated experiments. When only the original and the estimated sources are available, the system identification method is used to estimate the energy flow network before performing signal decomposition. The work in this paper is integrated in a uniform BSS system for research and real-word application purpose [14].

In BSS community, most evaluation algorithms require the original source signals are known in advance, however, this assumption cannot be satisfied in real-world applications. Only a few methods are proposed to evaluate the separation performance for real data, for example: in [22], the pitch information was extracted from the separated speeches to help the evaluation. This evaluation method, although useful in its special environments, is still not general enough for normal real-word evaluation purpose. How to evaluate the separation performance in real-world source separation applications is still an open problem needs to be further studied.

References

[1] S. Makino, H. Sawada, R. Mukai, S. Araki, Blind Source Separation of Convolutive Mixtures of Speech in Frequency Domain, IEICE Trans. Fundamentals, vol. E88-A, no. 7, pp. 1640-1655, (2005).

[2] Kun Gai, Zhenwei Shi, Changshui Zhang, Blind Separation of Superimposed Moving Images Using Image Statistics, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pp. 19-32, (2012).

[3] Yueyue Na, Jian Yu, Independent Vector Analysis Using Subband and Subspace Nonlinearity, EURASIP Journal on Advances in Signal Processing 2013, 2013: 74.

[4] E. Bingham and A. Hyvarinen, A Fast Fixed-point Algorithm for Independent Component Analysis of Complex Valued Signals, International Journal of Neural Systems, vol. 10, no. 1, pp. 1-8, (2000).

[5] M. S. Pedersen, J. Larsen, U. Kjems, L. C. Parra, A Survey of Convolutive Blind Source Separation Methods, Springer Handbook on Speech Processing and Speech Communication, (2007).

[6] Taesu Kim, Hagai T. Attias, Soo-Young Lee, Te-Won Lee, Blind Source Separation Exploiting Higher-Order Frequency Dependencies, IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 1, pp. 70-79, (2007).

[7] Intae Lee, Taesu Kim, Te-Won Lee, Fast Fixed-point Independent Vector Analysis Algorithms for Convolutive Blind Source Separation, Signal Processing, vol. 87, no. 8, pp. 1859-1871, (2007).

[8] H. Sawada, S. Araki, S. Makino, Underdetermined Convolutive Blind Source Separation via Frequency Bin-Wise Clustering and Permutation Alignment, IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 3, pp. 516-527, (2011).

[9] H. Sawada, S. Araki, S. Makino, Measuring Dependence of Bin-wise Separated Signals for Permutation Alginment in Frequency-domain BSS, IEEE International Symposium on Circuits and Systems, pp. 3247-3250, (2007).

[10] M. Z. Ikram and D. R. Morgan, Permutation Inconsistency in Blind Speech Separation: Investigation and Solutions, IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, pp. 1-13, (2005).

[11] E. Vincent, R. Gribonval, C. Fevotte, Performance Measurement in Blind Audio Source Separation, IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1462-1469, (2006).

[12] V. Emiya, E. Vincent, N. Harlander, V. Hohmann, Subjective and Objective Quality Assessment of Audio Source Separation, IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2046-2057, (2011).

[13] The bss_eval toolbox, http://bass-db.gforge.inria.fr/bss_eval/.

[14] The BSS platform, http://211.71.76.45/bss/ (or contact the first author directly).

[15] L. Wang, H. Ding, F. Yin, A Region-Growing Permutation Alignment Approach in Frequency-Domain Blind Source Separation of Speech Mixtures, IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 3, pp. 549-557, (2011).

[16] L. Rabiner, J. Allen, Short-Time Fourier Analysis Techniques for FIR System Identification and Power Spectrum Estimation, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-27, no. 2, pp. 182-192, (1979).

[17] J. Schoukens, G. Vandersteen, K. Barbe, R. Pintelon, Nonparametric Preprocessing in System Identification: A Powerful Tool, European Journal of Control, vol. 3-4, pp. 260-274, (2009).

[18] H. Sawadas dataset, http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html.

[19] J. Allen, D. Berkley, Image Method for Efficiently Simulating Small-room Acoustics, Journal Acoustic Society of America, vol. 65, no. 4, pp. 943-950, (1979).

[20] Yueyue Na, Jian Yu, Kernel and Spectral Methods for Solving the Permutation Problem in Frequency Domain BSS, International Joint Conference on Neural Networks, pp. 1-8, (2012).

[21] The cocktail party problem demo, http://research.ics.tkk.fi/ica/cocktail/cocktail_en.cgi.

[22] Yangfeng Liang, S. Naqvi, J. Chambers, Audio Video Based Fast Fixed-point Independent Vector Analysis for Multisource Separation in a Room Envionment, EURASIP Journal on Advances in Signal Processing, vol. 183, (2012).