ABSTRACT

Independent component analysis (ICA) is a kind of steady algorithm for the separation of blind source (BSS). However, there are three assumptions for the ICA application, one of which is that the number of test channels must be more than that of the signal sources. The limitation brings much inconvenience to practical signal acquisition and processing. In this paper, we propose a new method of repeated independent component analysis (Re-ICA) to realize the separation of blind sources. Under a single test channel, we can increase test channels by means of constructing virtual signal channels and use ICA repeatedly to separate every source signal in turn. Numerical simulation and signal processing of practical data acquisition by one single test-channel show that the proposed method is simple and feasible for operation, and is of great potential in engineering application.

1 INTRODUCTION

Blind source separation (BSS) is one of the attractive researches in signal processing [1-3]. Independent component analysis (ICA) is a special BSS method [4-6]. It is widely used in biomedical [6], machinery fault diagnosis [7], communication technology [8] and other areas. The processing object of the ICA is a group of mixed signals which are linearly composed of mutually statistically independent sources. The BSS purpose is to separate every independent component from the mixed-signal sources. There are three assumptions for the BSS application. One of the three assumptions is that the number of test channels must be more than that of the signal sources [1]. This limitation brings much inconvenience to the BSS for practical signal acquisition and processing. For example, it is hard to know the number of signal sources in practical test, and hence it will possibly cause waste for the measuring instrument and produce much difficulty for the experiment if we blindly increase the number of the test channels to identify the signal sources. If the number of the channels of signal acquisition is less than that of the signal sources, the credibility of the signal separation by ICA is more problematical. In this paper, we propose a new separation method which is named as repeated independent component analysis (Re-ICA). By constructing a virtual-channel signal, we can separate source signals based on spectrum identification and Re-ICA. The proposed method is very suitable for the detection of weak signals covered or suppressed by other large interferential signals. Particularly, we can recycle to use ICA to separate various weak source signals in turn.

2 THE PRINCIPLE OF REPEAT INDEPENDENT COMPONENT ANALYSIS

Suppose the original test data $\mathcal{X}(t)$ contain several signal sources which possess different strength. We use a single measurement channel to capture the mixed original data $\mathcal{X}(t)$. For separating the different signal sources, the processing steps of Re-ICA are as follows: 1) We analyze the collected data $\mathcal{X}(t)$ with FFT spectrum and identify the clear high spectral peaks. It means that we first identify part of the signal sources $\mathcal{S}_1(t)$. 2) According to the large spectral peaks, i.e., the identified signals in step one, we construct the additional virtual-channel signals that are corresponding to the identified signals. In this way, we can increase the number of the test channels without changing the signal sources. 3) We mix the original test data with the virtual-channel signals and separate the newly mixed signals by ICA. We can get two groups of split signals: one signal group with spectral peaks $\mathcal{S}_1(t)$, i.e.,
the identified signals in step one, and the other group 
with spectral peaks $S_2(t)$. $S_2(t)$ is some smaller source signals in the 
original test data $X(t)$. Here we define the data group involving 
the spectral peaks $S_2(t)$ as residual components $y_i(t)$. 
Repeating the previous steps 2) and 3) to process the residual 
components, we can detect much weaker signal sources. If 
continuing such a recycling procedure, eventually we can 
separate and identify all of the weak signal sources in the 
original test data. The mechanism of our method is shown in 
Fig.1.

For the case of a single channel test, suppose the tested 
original signal $s(t)$ only includes two signal sources, $s_1(t)$ 
and $s_2(t)$ , and $s(t)$ is expressed as $s(t) = [s_1(t), s_2(t)]^T$. 
Here $s_1(t)$ is the main component and we define it as a strong 
signal because of its larger energy in the original signal $s(t)$. 
Comparatively, $s_2(t)$ is a weak signal with small energy. 
Assume that the signal acquired or observed with the single 
channel test is $x(t) = as_1(t) + bs_2(t)$, and the parameters $a$ 
and $b$ are the amplitude weighting of $s_1(t)$ and $s_2(t)$ , 
respectively. Then the basic model of ICA is given by [1-7]

$$x(t) = As(t) = A \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = (a \ b) \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix}$$  \hspace{1cm} (1)

Where the mixture matrix is $A = (a \ b)$.

From the power spectrum of the observed signal $x(t)$, it is 
easy to detect the spectral component of $s_1(t)$, but the weak 
signal $s_2(t)$ may be too small to identify. According to the 
proposed method of Re-ICA, the principle steps of the 
analysis model are as follows:

1. We can identify the large signal $s_1(t)$ through the spectrum 
of $x(t)$. Then we construct a simulation sinusoidal 
signal $s_1'(t)$ which has the same frequency and phase as $s_1(t)$, 
and the data length of $s_1'(t)$ should be the same as that of the 
observation signal $x(t)$. The constructed signal $s_1'(t)$ is just 
the virtual-channel signal.

2. By means of a mixture matrix ($2 \times 2$), we mix the simulation 
signal $s_1'(t)$ with the original observation signal $x(t)$ to form a 
new observation signal $x'(t)$. So $x'(t)$ has one more channel 
than the original signal $x(t)$ to become a dual-channel 
observation signal.

3. Suppose this re-mixed matrix is described as $B$, it is a $2 \times 2$ 
square matrix. We can get the weak signal $s_2(t)$ using ICA to 
separate $x'(t)$. If the virtual signal $s_1'(t)$ is fully consistent 
with $s_1(t)$, i.e., $s_1'(t) = s_1(t)$, then the calculation model for 
the separation is written as

$$x'(t) = B \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = B \begin{pmatrix} s_1(t) \\ s_1'(t) \end{pmatrix} = B \begin{pmatrix} x_1(t) + bx_2(t) \\ s_1'(t) \end{pmatrix}$$

According to the ICA principle, if a mixing matrix is 
invertible, it would be beneficial to separation. Since the 
matrix $B$ is defined artificially and arbitrarily, so we can just 
define the matrix $B$ as an inverse matrix. For universality, let 
the matrix $B$ be the identity matrix $E$. Obviously, it is 
invertible. In practical significance, this means that, under an 
ideal experimental condition, only the ingredient $s_1(t)$ can be 
measured with an ideal sensor. Then, Eq.(2) can be written as

$$x'(t) = B \begin{pmatrix} a \\ 1 \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = E \begin{pmatrix} a \\ 1 \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = \begin{pmatrix} a \\ 1 \end{pmatrix} s_1(t)$$  \hspace{1cm} (3)

Let $A' = \begin{pmatrix} a \\ 1 \end{pmatrix}$ and $A'$ becomes a new mixed-matrix. So Eq. 
(2) is simplified as

$$x'(t) = A' \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix}$$  \hspace{1cm} (4)

Comparing Eq. (4) with the basic ICA model (1), we can find 
that the transformation from $A$ to $A'$ is a reversible 
transformation and that this transformation does not change 
the original signal $s(t)$. Adding the virtual channel implies 
increasing the number of the test channels of the observed 
signal. Therefore, the two models, Re- ICA and ICA, have the 
same form and property, and the Re- ICA model just needs 
more appropriate transformation than the ICA model, which 
makes $s_1(t)$ and $s_2(t)$ be refined and separated again.

In Eq.(4), if $s_1(t)$ and $s_2(t)$ respectively belong to the large-signal 
parts and the weak-signal parts in the original observed 
signal, we can use the proposed method repeatedly to process 
the weak-signal parts (residual components $y_i(t)$, ref. Fig.1) 
continuously until all the weaker signals are extracted.

### 3 SIMULATION EXPERIMENT

Suppose the simulated signal is a one-dimension signal $x(t)$ 
with 2000 data points and the numerical step is 0.0005s, i.e., 
sampling frequency is 2k Hz. The signal is arbitrarily 
composed of three sinusoidal signals whose frequencies are 
100 Hz, 90 Hz and 85 Hz respectively, and the corresponding 
amplitudes are 100, 1 and 0.001 in turn. These three signals 
are of different intensity. The amplitude of 100 Hz signal is the 
greatest. The intensity of 90 Hz signal is 1/100 times that
of 100 Hz signal, and the intensity of 85 Hz signal is 1/100 times that of 90 Hz signal and 1/10^4 times that of 100 Hz signal. Compared with 100 Hz signal, the signals of 90 Hz and 85 Hz are weak signals. Let

\[ x(t) = 100 \sin(2\pi \times 100t) + \sin(2\pi \times 90t) \\
+ 0.01 \sin(2\pi \times 85t) \]

(5)

Fig. 2 Time waveform and frequency spectrum.

Fig. 2 shows the time domain waveform and frequency spectrum of the original simulated signal \( x(t) \). From the spectrum, only the ingredient of 100 Hz signal can be recognized, while the other two frequency signals can not be extracted.

According to the proposed method of Re-ICA, we construct a virtual-channel signal \( S_i(t) = 100 \sin(2\pi \times 100t) \) as the large-signal, i.e., the recognized signal in Fig.2, and mix it with \( x(t) \) by means of a two-dimension identity matrix. Then we get a two-dimension signal \( x'(t) \). It is a 2×2000 matrix. We separate \( x'(t) \) through ICA and obtain its split frequency spectra shown in Fig.3 (a).

(a) Separation for setting 100 Hz virtual-channel signal.

(b) Separation for re-setting 90 Hz virtual-channel signal.

Fig. 3 The Re-ICA spectra of the data in Fig.2.

In Fig.3 (a), besides separating 100 Hz signal (the upper graph), we also detect the weak signal of 90 Hz frequency component (the bottom graph). Here we define the component containing the new separated 90 Hz signal as the residual component depicted as \( y_i(t) \), see Fig.1. To recognize the weakest 85 Hz signal in the residual component \( y_i(t) \), we repeat the above procedure to process \( y_i(t) \). That is, we construct the virtual-channel signal \( S_i(t) = \sin(2\pi \times 90t) \) according to the new separated signal in Fig.3 (a), and mix it with \( y_i(t) \) to form a two-dimension signal \( y'_i(t) \). Separating \( y'_i(t) \) by ICA, we extract the weakest 85 Hz signal shown in the bottom spectral graph of Fig.3 (b). The upper spectral graph in Fig.3 (b) is the 90 Hz signal component which is separated again.

4 DISCUSSIONS

In the Re-ICA process, the value of the amplitude for constructing a virtual-channel signal has little influence on the separation result because the signal amplitude in ICA is uncertainty [1, 4]. That is, no matter what the amplitude of the virtual-channel signal is, the spectral amplitude of the Re-ICA separation is always a constant. However, if the amplitude of the virtual-channel signal is set too small, such as the ratio of virtual amplitude and the real one much smaller than 1, then the Re-ICA separation may not be available. This indicates that if the amplitude difference between the virtual-channel signal and the detected large-signal is too great, the Re-ICA will be failure. To avoid this case, we can take the amplitude of the virtual-channel signal close to the size of the amplitude of the detected large-signal. The amplitude of the detected large-signal can be approximately valued from its FFT spectrum.

When we determine the frequency of a virtual-channel signal in terms of the method of Re-ICA, if the frequency has an error, that is, the determined frequency is not exactly equal to the real frequency because of, for example, inaccurate observation or different frequency resolution, this inaccuracy will affect the Re-ICA effect. Suppose the determined signal for simulation is \( x(t) = 100 \sin(2\pi \times 100t) + \sin(2\pi \times 90t) \) and data points for calculation are 2000 as well as the numerical step is 0.0005s (frequency sampling is 2k Hz). To separate the weak 90 Hz frequency signal \( \sin(2\pi \times 90t) \), the virtual-channel signal should be \( 100 \sin(2\pi \times 100.2t) \). However, because the frequency resolution is 1 Hz, the real frequency 100.2 Hz cannot be accurately distinguished and only the optimal frequency 100 Hz can be evaluated to approximate to 100.2 Hz. So if the frequency of the virtual-channel signal is read as 100 Hz, that is, the frequency error between the virtual-channel signal and the FFT detected large-signal is \( \Delta f = 0.2 \) Hz, then the Re-ICA result is shown in Fig.4 (a). The weak 90 Hz frequency signal cannot be extracted.

If the frequency of the virtual-channel signal is improved as 100.19 Hz (or 100.21 Hz) closer to the real frequency 100.2 Hz, the Re-ICA result is depicted in Fig.4 (b). Compared with the case of Fig.4 (a), the spectral peak of the weak 90 Hz signal becomes to emerge, note the bottom graph of Fig.4 (b). But the 90 Hz frequency signal is still covered by the large 100 Hz frequency signal in spite of the smaller frequency error \( \Delta f = 0.01 \) Hz. When the frequency of the virtual-channel
signal is further close to the real frequency, for instance, let it be 100.199 Hz (or 100.201 Hz), the weak signal of 90 Hz frequency can be detected more obviously in the Re-ICA result presented in Fig.4 (c).

In fact, it is often inevitable to construct an inexact frequency of a virtual-channel signal because the virtual-channel signal frequency is hard to coincide with the source signal frequency in practice. This will cause imperfect correlation between the virtual-channel signal and the source signal. As a result, the residual component of the separated signal group still contains part of the virtualized source signal. Therefore, to make the virtual-channel signal frequency much closer to or equal to the real source signal frequency and to improve the accuracy of the Re-ICA method, the frequency of the virtual-channel signal can be determined by searching the real source frequency around the frequency of the large FFT spectral peak in different frequency resolution.

As for the phase, in practical application, it is difficult to make the phase of the virtual-signal and the real signal much closer or equal and to improve the accuracy of the Re-BSS, according to the frequency processing way, we can determine the phase of the virtual-channel signal by searching the real phase of the large source signal in different phase resolution.

5 ENGINEERING APPLICATION ---- THE FAULT DIAGNOSIS OF A ROTOR SYSTEM

In a small rotor system, the rotating speed of the rotor is 375 rpm or the running frequency of the rotor is 6.25 Hz. One terminal section of the rotor is a hexagon shape. The hexagon end is enclosed in a circular iron barrel. In the normal working condition, there is an interspace between hexagon end and the iron barrel. If the rotor is not balanced well or out of alignment, the hexagon end may rub-impact inner wall of the circular barrel when the rotor runs. To monitor the working condition of the rotor and to determine whether the rotor contacts the barrel, a piezoelectricity acceleration sensor is fixed at outer wall of the barrel and the signal from the sensor is real-timely sampled with NI Express Card-8360 devices. The sampled data is finally recorded in a computer which connects with the NI card. Fig.5 shows the monitoring result in one test. The sampling frequency is 10 kHz and the data length from the recorded data for FFT spectrum calculation is 6500 points.

As is shown in Fig.5, the larger frequency components lie at 37.5 Hz and its harmonic frequency 70 Hz. The frequency 37.5 Hz is not the rotor’s running frequency 6.25 Hz. It is six times the running frequency of the rotor. By our proposed Re-BSS method and through searching the real frequency around the frequencies of 37.5 Hz and 75 Hz as well as searching the real phase, we determine $2 \sin(2\pi \times 37.45t + \pi)$ and $\sin(2\pi \times 74.9t)$ as the two virtual-channel signals. The recognition results are shown in Fig.6.

Figures in Fig.6 (a) are the time waveforms and the corresponding FFT spectra are presented in Fig.6 (b). The bottom graph in Fig.6 (b) shows the spectral features of the weak signals after the large signals at the frequencies of 37.5 Hz and 75 Hz are significantly weakened. These detected weak spectral spikes contain the components of running frequency 6.25 Hz and its harmonics. The Re-BSS spectral spikes at the running frequency and its harmonics may forecast early and weak mechanical faults in the rotor system, such as rub-impact of the rotor, shaft-misalignment, and rotor imbalance.

Since in the structural design of this rotor system, the fault of shaft-misalignment is excluded, the only source of faults comes from the rotor imbalance and rub-impact in the rotor system. Therefore, we conclude that the characteristic spectral peak at the running frequency corresponds to the weak imbalance of the rotor, and that this weak rotor imbalance resulted in the rub-impact and formed the large spectral peak at frequency 37.5 Hz owing to six-time collision of hexagon end when the rotor rounds a single circle.
6 CONCLUSIONS
In the issue of BSS based on ICA, one of the restrictive conditions is that the number of the test channel is not less than that of the signal source. To improve the test channel of the BSS, we propose the method of repeated ICA in this paper. Our method can break through the limit of the number of test channel by constructing virtual-channel signals. Numerical simulation of one single channel and signal processing of practical data acquisition show that the proposed method can be manipulated easily, and is of great potential in engineering application.

7 ACKNOWLEDGEMENT
This work is supported by the National Natural Science Foundation of China (NSFC, granted No. 50975202).

8 REFERENCES