REGION-GROWING PERMUTATION ALIGNMENT APPROACH IN FREQUENCY-DOMAIN CONVOLUTIVE BLIND SOURCE SEPARATION

Lin Wang[1,2], Heping Ding[1], and Fuliang Yin[2]
[1] Institute for Microstructural Sciences, National Research Council Canada, lin.wang@nrc-cnrc.gc.ca
[2] School of Electronic and Information Engineering, Dalian University of Technology, P. R. China, 116023

1. INTRODUCTION

Blind source separation (BSS) is a technique for recovering the original source signals from observed signals with the mixing process unknown. One well recognized BSS application is the separation of audio sources that have been mixed and recorded by a microphone array in a real room environment, i.e. a cocktail party problem. The challenge of the problem is that the mixing process is convolutive, where observations are the combinations of the unknown filtered versions of the sources. Frequency-domain BSS methods are promising for convolutive BSS due to its fast convergence and small computation load[1].

In frequency-domain BSS, the convolutive separation problem is converted to an instantaneous Separation problem in each frequency bin by short-time Fourier transform (STFT). Generally, satisfactory instantaneous separation may be achieved within all frequency bins, but combining them to recover the original sources is a challenge because of the unknown permutations associated with individual frequency bins. The permutation ambiguities should be looked after properly so that the separated frequency components from the same source are grouped together.

To solve the permutation problem in frequency-domain blind source separation, this paper proposes a new alignment method based on an inter-frequency dependence measure: the powers of separated signals. Bin-wise permutation alignment is applied first across all frequency bins, using the correlation of separated signal powers; then the full frequency band is partitioned into small regions based on the bin-wise permutation alignment result; and finally, region-wise permutation alignment is performed in a region-growing-like manner. Experiment results verify the effectiveness of the proposed method.

2. METHOD

2.1 Frequency-domain BSS

The convolutive separation problem can be converted to instantaneous separation in each frequency bin via short-time Fourier transform (STFT). In the frequency domain, it is possible to separate each frequency bin independently using complex-value instantaneous BSS algorithms such as Infomax[2], which are considered as quite mature. However, there is a scaling and permutation ambiguity in each bin, where the order and scale of the separated signals are unknown. It is necessary to correct the scaling and permutation ambiguity before transforming the signals back to the time domain. The permutation correction is a challenging problem. The permutation at each bin should be aligned so that the separated components originating from the same source are grouped together. The scaling ambiguity can be relatively easily resolved by using the Minimal Distortion Principle[3]. Finally the unmixing network $W(n)$ is obtained by inverse discrete Fourier transforming (IDFT) and the estimated source $y(n)$ is obtained by filtering $x(n)$ through $W(n)$. The workflow of the frequency-domain BSS is shown in Fig. 1.

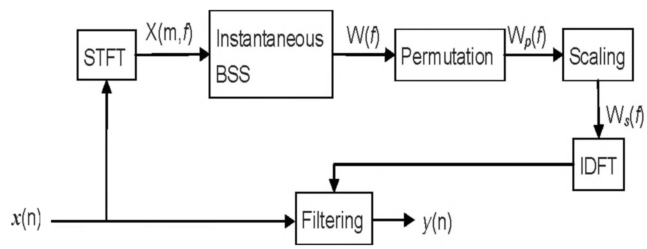


Fig. 1 Workflow of frequency-domain blind source separation

2.2 Proposed permutation alignment scheme

2.2.1 Inter-frequency dependence measure

The inter-frequency dependences of speech sources can be exploited to align the permutations across all frequency bins. An inter-frequency dependence measure proposed in [4], the correlation coefficient of separated signal power ratios, exhibits a clearer inter-frequency dependence among all frequencies. Its definition is

$$v_i(f) = \frac{\text{power of the } i\text{'th separated signal at } f}{\text{total power of all the separated signal at } f} \quad (1)$$

The power ratio represents the dominance of the i'th separated signal in the observations at frequency f. Being in the range [0, 1], (1) is close to 1 when the i'th separated signal is dominant, and close to 0 when others are dominant. The power ratio measure can clearly exhibit the signal activity due to the sparseness of speech signals.

2.2.2 Permutation alignment scheme

The correlation of bin-wise signal power ratio tends to be high when the two components belong to the same source, but such dependence is not always evident. Generally the misalignment at an isolated frequency bin may spread to other frequencies easily and cause a big misalignment beyond it. A region-growing-like permutation scheme is proposed which can survive this misalignment spread problem. The scheme is described in 4 steps as follows.

Step 1. Calculate the power ratio (1) for all frequency bins and all separated signals

Step 2. Correct permutation bin by bin, so that the power ratio time sequences at each frequency bin has the highest correlation coefficient with the previous bin.

Step 3. Partition the frequency band into K regions, where the frequency bins with high correlations belong to one region.

Step 4. Select a region with the largest number of elements as a seed; merge with its neighboring regions on both sides in a region-growing style.

Based on the description above, the workflow of the proposed scheme is shown in Fig. 2.

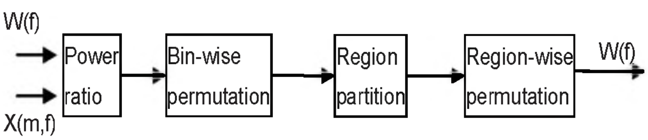


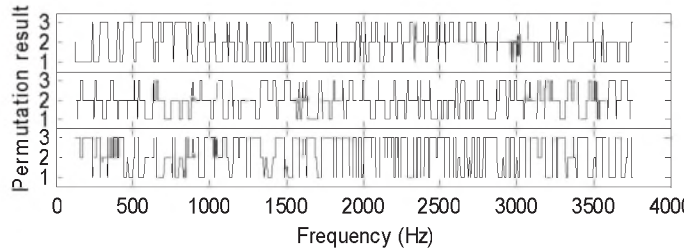
Fig. 2 Workflow of proposed permutation alignment method

3. RESULTS

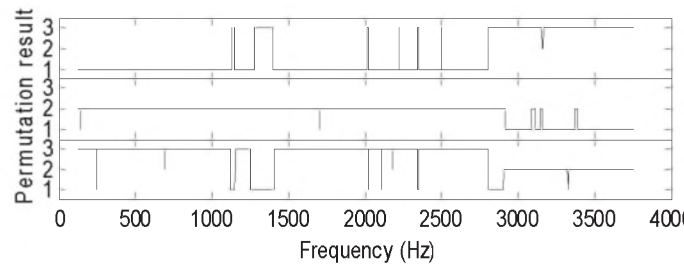
In this experiment the proposed algorithm is applied to the problem with three microphones and three sources in a simulated environment. The room size is $7\text{m} \times 5\text{m} \times 3\text{m}$, the simulated room reverberation time is RT60 = 130 ms, where RT60 is the time required for the sound level to decrease by 60 dB. To demonstrate the permutation alignment performance, we show the permutation result at three stages in Fig. 3: (a) before permutation alignment, (b) after bin-wise permutation alignment, and (c) after region-wise permutation alignment. The permutation result is calculated using the method proposed in [5], supposing the mixing filters are known. It can be seen from Fig. 3 that the permutation ambiguity is very severe before permutation alignment; the ambiguity is mitigated after bin-wise permutation alignment but large misalignments still occur; and the ambiguity is almost eliminated after region-wise permutation alignment except some misalignments on some isolated frequency bins. Such misalignments do not spread to nearby frequencies. Finally, the average output signal-to-interference ratio is 14.2 dB, a 16.8 dB improvement over the average input signal-to-interference ratio of -2.6 dB.

4. CONCLUSION

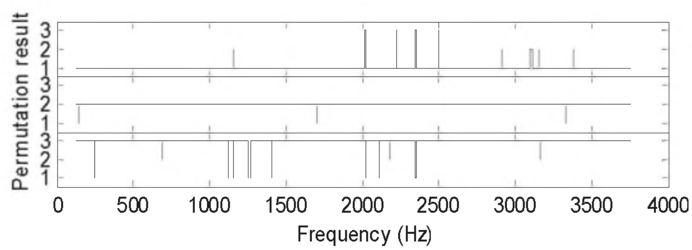
Studying frequency-domain convolutive blind source separation, this paper proposes a new permutation alignment method based on the inter-frequency dependence of the correlation of separated signal powers. With a region-growing-like permutation alignment style, the proposed method minimizes the spreading of misalignment at isolated frequency bins to others. Good separation results are observed in simulation.



(a) Before permutation alignment



(b) After bin-wise permutation alignment



(c) After region-wise permutation alignment

Fig. 3 Permutation result of the proposed method

REFERENCES

[1] Sawada, H. (2007). Frequency-domain blind source separation, in Blind Speech Separation, 47-78, Springer.
[2] Douglas, S.C. (2007). Scaled natural gradient algorithms for instantaneous and convolutive blind source separation, ICASSP 2007, vol. 2, 637-640.
[3] Matsuoka, K. (2001). Minimal distortion principle for blind source separation, 2001 International Workshop on Independent Component, 722-727.
[4] Sawada, H. (2007). Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS, 2007 IEEE International Symposium on Circuits and Systems, 3247-3250.
[5] Ikram, M.Z. (2000). Exploring permutation inconsistency in blind separation of speechsignals in a reverberant environment, ICASSP 2000, vol. 2, 1041-1044.

ACKNOWLEDGEMENT

This work was supported by NRC-MOE Research and Post doctoral Fellowship Program from Ministry of Education of China and National Research Council Canada.