EFFICIENT BLIND SPEECH SIGNAL SEPARATION COMBINING INDEPENDENT COMPONENT ANALYSIS AND BEAMFORMING

Qiongfeng Pan and Tyseer Aboulnasr
School of Information Technology and Engineering, University of Ottawa, 800 King Edward Ave., Ottawa ON, K16 6N5
aboulnasr@eng.uottawa.ca

1. Introduction

Teleconferencing is a common application where we need to separate different speakers whose speech is picked up by any given microphone. Such a problem has been tackled using Blind Source Separation (BSS) algorithms utilizing time and frequency domain information [1] and beamforming (BF) algorithms utilizing spatial information [2] from different point of views.

We investigate ways to combine these two approaches for improved overall system performance. While the popular BF approach utilizes the spatial information about the mixing system and/or source signals, BSS exploits a strong statistical condition: independence between source signals. These approaches have much in common since source signals coming from different locations in a BF scenario are likely to be independent as well. As such, it is worthwhile to explore the possibilities of combining their advantages. Recently, the relationship between convolutive BSS and BF has been investigated in [3] and some interesting results have been obtained. Based on these results, some combinations of convolutive BSS and BF have been proposed [7] to solve problems in BSS and obtained improved separation results.

In this paper, we propose a new approach combining BSS and BF for blind speech signal separation in real acoustic environment building on the work in [7]. In the beamforming stage, the Directions of Arrival (DOA)s of sources of interest are estimated blindly; then beamformers are constructed to extract signals from these directions. In the BSS stage, frequency domain convolutive algorithm is utilized to further reduce the interference in the given direction and improve the separation performance. Compared with existing systems, the proposed approach significantly reduces the computational complexity while maintaining comparable separation performance.

2. Convolutive Blind Source Separation and Beamforming

The convolutive BSS model is illustrated in Fig. 1. N source signals $\{s_i(k)\}$, $1 \le i \le N$, pass through an unknown N-input, M-output linear time-invariant mixing system to yield the M mixed signals $\{x_j(k)\}$. All source signals $s_i(k)$ are assumed to be statistically independent.

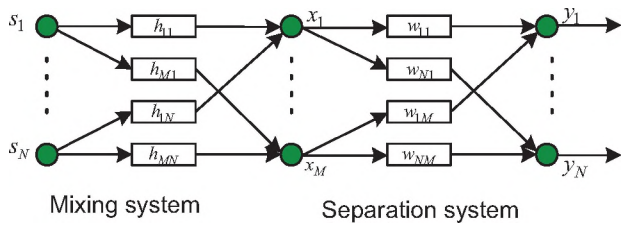

Fig. 1: Structure of convolutive blind source separation system

The j th sensor signal can be obtained by

$$x_j(k) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} h_{ji}(l) s_i(k-l) \qquad (1)$$

where $h_{ji}(l)$ is the impulse response from source i to sensor j, L defines the order of the FIR filters used to model this impulse response.
The i-th output of the unmixing system is given as:

$$y_i(k) = \sum_{j=1}^{M} \sum_{l=0}^{Q-1} w_{ij}(l) x_j(k-l) \qquad (2)$$

By extending the instantaneous BSS algorithm to the convolutive case, we get the time domain update for the convolutive BSS algorithm using the natural gradient optimization approach [6]:

$$\Delta \underline{\mathbf{W}} = -\mu \frac{\partial D}{\partial \underline{\mathbf{W}}} \underline{\mathbf{W}}^T \underline{\mathbf{W}} = \mu \left[\mathbf{I} - E\left(\varphi(\mathbf{y}) \mathbf{y}^T \right) \right] \underline{\mathbf{W}} \qquad (3)$$

where $\underline{\mathbf{W}}$ the unmixing matrix with FIR filters as its components. Convolutive BSS can also be performed in the frequency domain by using short-time Fourier Transform. This method is based on transforming the convolutive blind source separation problem into an instantaneous BSS problem at every frequency bin.

In [7], independent component analysis (ICA) is used to perform blind source separation at every frequency bin and the unmixing matrix obtained. Accordingly, the directivity pattern at each frequency bin can be obtained from its unmixing matrix. Directions of arrival (DOA) of source signals are estimated from the directions of nulls at all frequency bins. In the adaptation process, at each frequency bin, the null direction in the directivity pattern is compared with the estimated DOA of source signals. If it is steering to the proper direction, the unmixing matrix from ICA algorithm is used. If not, the null-steering beamformer constructed from the estimated DOA information is used to

substitute for the unmixing matrix. By doing so, the unmixing matrix can recover from a local minimum in the optimization procedure to improve its convergence speed.

At every iteration and at each frequency, the ICA algorithm is used to update the weight coefficients; then the DOA information at this frequency is obtained by searching for the null from the directivity pattern of unmixing matrix; the beamformer is formed for every frequency and a comparison is conducted between ICA and beamformer directions. All these operations are very time consuming and require significant computation. The separation performance is also very sensitive to the frame length of the frequency domain BSS algorithm and requires large frame lengths further increasing the computational complexity.

Studying the approach in [7], we can show that: i) low and high frequency bands do not provide good estimations of the DOA, ii) the accuracy of estimated DOAs is effectively independent of frame length, iii) A subset of frequency bands is sufficient to determine the unmixing matrix at every frequency bin to obtain DOA estimation.

3. Proposed Combined BF- BSS System and Simulation Results

The proposed system has two stages: i) blind BF used to obtain signal from estimated source directions and reduce reverberation effects ii) frequency domain BSS to further separate residual interferences in the selected direction and improve the separation performance.

In the BF stage, we implement a new DOA estimation algorithm [5] where the DOA can be independently estimated for mid frequency bins only effectively reducing the computational complexity. In the convolutive BSS stage, an unmixing system **W** is adaptively adjusted to make the outputs as independent as possible to recover the independent source signals.

The mixed signals are generated by convolving speech signals with measured real room impulse responses. One signal is located at a DOA of 20 degrees and the other one is at a DOA of 60 degrees. The PESQ (Perceptual Evaluation of Speech Quality) score [4] is used to measure the subjective quality of the recovered speech signal compared to the original speech signal at each stage. The PESQ scores for the mixed signals, output signals from BF stage and output signals from BSS stage compared with the original signals are shown in Table 1. For the original mixture, each mixed speech signal is almost equally similar to both sources. BF and BSS stage ensure each output signals is more biased to one source and away from the other. This was confirmed by our informal listening test.
Since the reverberant effects have been already reduced by the beamforming stage, the frame size of the FFT is much smaller than that used in [7]. Thus, the computational complexity is further reduced.

Table 2 shows the PESQ improvement (defined as the sum of the PESQ score away from one source and PESQ score bias to the other source). Table 2 shows the that the proposed system performs very well for speech separation in a real acoustic environment with reduced complexity and flexible system structure.

PESQ	Mixtures		Outputs from BF stage		Outputs from BSS stage	
	x1	x2	y1	y2	z1	z2
s1	1.62	1.60	1.73	0.63	2.11	0.25
s2	1.47	1.50	1.07	2.16	0.59	2.32

Table 1: PESQ scores from different stages

PESQ Improvement	BF Stage		BSS Stage	
s1	0.26	0.87	0.38	0.38
s2	0.55	0.56	0.48	0.16
Total	0.81	1.43	0.86	0.54

Table 2: PESQ improvement from different stages

4. Conclusions

By utilizing the properties of the unmixing matrix in the freq. domain, we propose a reduced complexity combined BF-BSS algorithm for speech separation in real acoustic environment. Performance is confirmed using PESQ scores.

REFERENCES

[1] S. Haykin, ed., Unsupervised adaptive filtering (Volume I: Blind Source Separation), John Wiley & Sons, 2000.

[2] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag., Apr. 1988, pp.4-24.

[3] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures, " EURASIP Journal on Applied Signal Processing, vol. 2003, no. 11, pp. 1157-1166, Nov. 2003.

[4] ITU-T Recommend P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to end speech quality assessment of narrowband telephone network and speech codecs," May 2000.

[5] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation, " IEEE Trans. Speech Audio Processing, vol. 12, no. 5, pp. 530-538, Sept. 2004.

[6] S. Amari and A. Cichocki, "A new learning algorithm for blind signal separation," In advances in neural information processing systems 8, pp.757-763, MIT press, 1996.

[7] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," IEEE Trans. Speech Audio Processing, Mar. 2006.