COMBINED SPATIAL/BEAMFORMING AND TIME/FREQUENCY PROCESSING FOR BLIND SOURCE SEPARATION

Qiongfeng Pan and Tyseer Aboulnasr

School of Information Technology and Engineering, University of Ottawa
800 King Edward Avenue, K1N 6N5, Ottawa, Canada
email: qpan@site.uottawa.ca; aboulnasr@eng.uottawa.ca

ABSTRACT

Convolutive blind source separation and adaptive beamforming have similar goals and similar system structure. Both attempt to extract selected source signals from observed sensor mixtures by a filter array. However, time and frequency information are utilized in convolutive blind source separation while spatial information of source signals or sensor array is used in adaptive beamforming. In this paper, we start with a brief introduction of blind source separation and adaptive beamforming. Next, we review approaches combining spatial information used in beamforming with time/frequency processing used in convolutive blind source separation. We also present a new proposed combination approach and simulation results.

1. INTRODUCTION

In many applications, there is a definite need to recover signals that have been mixed together e.g. teleconferencing. Such a problem has been tackled using Blind Source Separation algorithms (utilizing time and frequency domain information) and adaptive beamforming (utilizing spatial information) from different points of views. In this paper, we review both approaches and investigate how to combine them for better system performance.

Independent component analysis (ICA) serves as a major statistical tool for solving the Blind Source Separation (BSS) problem. Separation is performed using the assumption that the source signals are independent with no information about the geometry of the auditory scene (such as direction of arrival of source signals, microphone array configuration etc.). Only time/frequency information of sensor signals are utilized in separation algorithms. However, some aspects limit further applications of BSS in real-world acoustic environments. These include low convergence rate and high computational requirements in time domain methods, frequency permutation and arbitrary amplitude scaling in frequency domain methods and performance degradation in heavy reverberant environments.

On the other hand, a relatively well-established research topic – adaptive beamforming for acoustical signals – approaches this problem from a spatial point of view. In adaptive beamforming, a structured array of sensors is used to steer the overall gain pattern of the array sensors to form a spatial filter which can extract signal from a specific direction and reduce signals from other directions. This enhances the receiver's performance with regards to source identifiability, direction tracking and quality of reception. Thus, compared with blind source separation, the advantage of adaptive beamfoming is that the available spatial information about the mixing system and/or source signals is utilized. However, blind source separation exploits a strong statistical condition -- independence -- between source signals, which should be helpful for adaptive beamforming.

Recently, the relationship between blind source separation and adaptive beamforming has been investigated in [2][7] [12] and some interesting results have been obtained. Based on these results, some combinations of blind source separation and adaptive beamforming have been proposed to improve separation results. In this paper, we analyse the relationship between blind source separation and adaptive beamforming in detail, review current combined approaches of blind source separation and beamforming, propose a new approach and present its simulation results.

2. CONVOLUTIVE BLIND SOURCE SEPARATION

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.

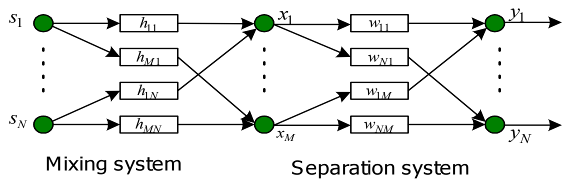

Fig. 1 Convolutive BSS model

The j th sensor signal can be represented as

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

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 the impulse response.

The task of the convolutive BSS is to obtain an unmixing system such that the outputs of this system $\mathbf{y}(k) = [y_1(k) \cdots y_N(k)]^T$ become mutually independent as estimates of the N source signals. The separation system typically consists of a set of FIR filters $w_{ij}(k)$ of length Q. The i th output of the unmixing system is given as:

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

In the time domain convolutive blind source separation algorithm, ICA is applied directly to the convolutive mixture model [17][18] resulting in good separation for small mixing filters. It is computationally very expensive for long FIR filters since it involves the convolution operation.

The frequency domain convolutive BSS algorithm is very popular for dealing with convolutive mixtures since the convolutive BSS problem is transformed into instantaneous BSS problem at every frequency bin. Any complex-valued instantaneous ICA algorithm can then be employed to deal with the separation at individual frequency bins. However, the permutation and scaling ambiguities are introduced *independently* at every frequency bin. This constitutes a major challenge that limits the potential application of frequency domain BSS as the components at different frequency bin may not come from the same source signal and may not have the same scale, creating a major problem in the time domain reconstruction of the signals.

3. ADAPTIVE BEAMFORMING

Typically, a beamformer [19] linearly combines the spatially sampled time series from each sensor to obtain a scalar output time series of a signal from a given direction, in the same manner as an FIR filter linearly combines temporally sampled data to select a signal in a given frequency range. The basic beamformer structure is given in Fig. 2.

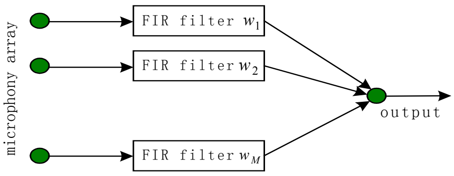

Fig. 2: Beamformer with senor outputs convolved by FIR filters

The filter coefficients are optimized to produce a spatial pattern with a dominant response in the direction of interest while the response for the directions of interfering signals is minimized. In multipath or reverberant environments, the selected direction may also include signals that originate from different sources that end up arriving in the same direction.

4. CURRENT APPROACHES FOR COMBINING BEAMFOMING AND BSS

It is clear that convolutive BSS and adaptive beamforming solve the same signal separation problem from different points of view using different information. To achieve better separation, combined BSS/beamforming approaches have been proposed along three general directions to be reviewed here.

4.1 Incorporation of geometrical information into convolutive blind source separation algorithm

In this first approach, geometric information, such as location of source (or direction of arrival) and sensor configuration used in beamforming to align the beam pattern to specific direction, is incorporated into the convolutive BSS algorithms.

In [12], the geometric information used in adaptive beamforming is incorporated in frequency domain convolutive BSS algorithm as linear constraints or as the initial adaptation condition. These additional constraints inevitably reduce existing degrees of freedom so as to resolve some of the ambiguities in convolutive BSS algorithm. It should be noted that an accurate steering direction is assumed to be known in [12]. This assumption is not always true. A new geometrically constrained BSS algorithm is proposed in [8] without this assumption. This algorithm is based on the FastICA algorithm [5] and roughly-estimated geometric information. The performance of this algorithm is not sensitive to the precision of the estimated geometrical constraint resulting in robustness of the algorithm in reverberant acoustical environment.

Besides incorporating geometrical information into the frequency domain convolutive BSS, a new time domain convolutive BSS algorithm is proposed in [1] by utilizing geometric information. In this algorithm, a null beamformer is constructed based on the available geometric information and its parameters are exploited as initial condition into the time domain convolutive BSS algorithm to speed its convergence rate and improve separation performance since the convergence and result of separation of gradient-based algorithms are influenced significantly by the initial conditions.

4.2 Formulation of convolutive BSS as multiple sets of adaptive beamforming to resolve ambiguities in BSS

A convolutive blind source separation system can be viewed as multiple sets of adaptive beamforming, which means the separation filter array for every output can be viewed as a beamformer. In [9], this idea is used to deal with frequency permutation and the arbitrary scaling problem in frequency domain convolutive BSS. After obtaining the unmixing matrix for every frequency bin, its corresponding directivity pattern can be calculated by beamforming approach. The null direction for every output at each frequency bin can be obtained from the directivity pattern. By swapping the output order of every frequency bin to make the output signals from frequency components consistent with the null direction, the frequency permutation problem can be resolved.

The directivity patterns used in [9] have grating lobes at high frequencies, which affect the accuracy of estimated direction of sources. In [6], the directivity patterns at different frequencies are investigated and a new approach is proposed by estimating the source location from the lower band of frequencies where no grating lobes appear. The frequency permutation is aligned by looking for nulls in the neighborhood of the estimated direction of arrival (DOA).

The directivity patterns obtained from unmixing matrix can also be used to improve the convergence speed of

convolutive BSS algorithm. In [15], independent component analysis and beamforming are combined to deal with the slow convergence problem. First, ICA is used to perform blind source separation at every frequency bin and the unmixing matrix can be obtained at each frequency bin. Accordingly, the directivity pattern at each frequency bin can be calculated from its unmixing matrix as in [9]. DOA of source signals are estimated from the directions of nulls at all frequency bins. During the adaptation process, at each frequency bin, the null 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 be recovered from local minimum in the optimization procedure to improve its convergence speed.
The approach proposed in [9] requires plotting the directivity pattern for every frequency bin; something that is very time consuming. Moreover, for situations with more than two sources, it is difficult to estimate DOA of source signals from null directions since the directivity pattern becomes too complicated. In [13], a new method dealing with the permutation problem in situations with more than two sources is proposed. In this approach, a closed-form formula is proposed to directly calculate direction of sources from the unmixing matrix at each frequency bin. By sorting the obtained directions of sources, a permutation matrix can be constructed to resolve the frequency permutation problem.
In [14], a new robust and precise method for solving frequency permutation in frequency domain convolutive BSS is proposed by integrating direction of arrival approach and interfrequency correlation approach [11]. Interfrequency correlation approach for frequency permutation alignment is based on the idea that signal envelops have high correlations at neighboring frequencies if separated signals are from the same source signal. However, the correlation approach is not robust since a misalignment at a given frequency can cause misalignments in subsequent frequencies. In this new method, for the frequencies where the direction of arrival can be estimated accurately, direction of arrival approach is used to align the frequency permutation, for other frequencies, interfrequency correlation approach is used to do the alignment based on neighbouring correlation.

4.3 Utilization of the beamforming structure and the ICA cost function

In [4], a new convolutive blind source separation algorithm is proposed based on a beamforming structure and the ICA cost function. In this method, the unmixing system is constructed as multiple sets of beamformers. Besides making null-steering towards interfering signal as in the conventional beamforming, the null-directions are also adjusted to make output signals as independent as possible. This means that the multiple sets of beamformers are adjusted jointly to obtain mutually independent outputs. Thus, the independence criterion, which includes higher-order statistics, and geometric information are both exploited in this algorithm.

5. A NEW COMBINATION OF CONVOLUTIVE BSS AND ADAPTIVE BEAMFORMING

The proposed system attempts to mimic the performance of human ears in a cocktail party environment. First, adaptive beamforming is used to isolate signals from specific directions, and then blind source separation is used to separate signal from different sources aiming in that direction. The proposed two-stage separation system is shown in Fig.3.

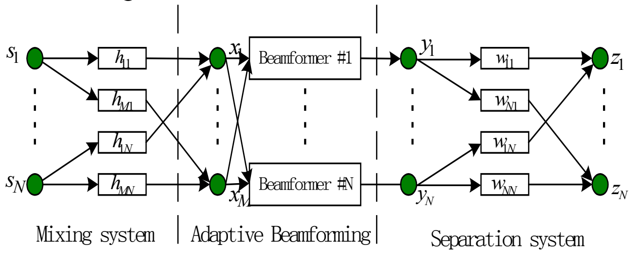

Fig.3 Proposed 2-stage beamforming BSS system of speech signals

The sound source localization and separation system proposed in [3] is implemented here as the adaptive beamforming stage to zoom in on the direction of selected speaker. First the range and direction of the speaker are estimated by an extended spatial spectrum estimator, MUSIC [16] for each source. Then the minimum variance beamformer is constructed based on the estimated location information.
In convolutive blind source separation stage, the update equation for the separating system $\underline{\mathbf{W}}$ is the algorithm used in [10] at frequency domain.
The two-stage setup mimicking the human hearing system allows the advantages of both adaptive beamforming and convolutive BSS to be implemented entirely with the freedom to select the best algorithm for each stage.

6. EXPERIMENTAL RESULTS

A microphone array with 8 sensors is used to receive 2 source speech signals sampled at 8kHz. Since the end users are humans, we use PESQ score [20] to measure the quality of the recovered speech signal as perceived by a human. The PESQ score is rated as a value between –0.5 to 4.5. The higher the score, the better the speech quality.
Real room impulse responses with 4096 taps are used to generate the 8 mixed signals $x_1, x_2, \cdots, x_8$. In the adaptive beamforming stage, the two speaker locations are estimated from the mixtures and two beamformers are constructed to get signals y_1 and y_2 from these two specific locations. In the convolutive BSS stage, the coefficients of the unmixing system are adaptively adjusted to further cancel the remaining effects of cross-talk and get the estimated source signals z_1 and z_2.
Tables 1, 2 and 3 provide the PESQ scores of the mixed signals $x_1, x_2, \cdots, x_8$; signals y_1 and y_2 from beamforming stage; signals z_1 and z_2 from BSS stage compared to the original signal, respectively.

PESQ		x1	x2	x3	x4	x5	x6	x7	x8
female/female	s1	1.831	1.811	1.798	1.746	1.600	1.663	1.559	1.739
	s2	1.462	1.260	1.306	1.336	1.342	1.559	1.662	1.488
male/male	s1	2.082	2.045	2.134	2.102	1.952	1.932	1.860	2.072
	s2	1.707	1.602	1.653	1.720	1.717	1.852	1.898	1.763
female/male	s1	1.842	1.854	1.826	1.742	1.621	1.668	1.528	1.781
	s2	1.537	1.428	1.498	1.565	1.616	1.738	1.802	1.529

Table 1:Average PESQ scores for mixed speech signals x_i

PESQ	female/female		male/male		female/male	
	y1	y2	y1	y2	y1	y2
s1	2.343	0.859	2.490	1.478	2.319	0.809
s2	0.665	2.154	1.091	2.298	0.752	2.292

Table 2: Average PESQ scores for outputs y_1 and y_2

PESQ	female/female		male/male		female/male	
	z1	z2	z1	z2	z1	z2
s1	2.411	0.680	2.672	1.039	2.460	0.550
s2	0.481	2.244	0.719	2.471	0.590	2.410

Table 3: PESQ scores for outputs z_1 and z_2

Since most of the reverberation effects have already been removed by the adaptive beamforming stage, we used very short FIR filters (32 taps) to complete the speech separation in the selected direction. Similar separation quality cannot be obtained even by filters with 1024 taps when there is no adaptive beamforming stage as pre-processor. Thus, the computation complexity for BSS is greatly reduced.
We can see that the combined beamforming and convolutive BSS algorithm further improves the quality of separation. This was also confirmed by informal listening experiments.

7. CONCLUSIONS

In this paper, we review approaches combining spatial information used in beamforming with time/frequency processing used in convolutive blind source separation aiming for better separation performance given the increased information used. We also present a new proposed combination method which mimics the way our ears separate audio signal in acoustic environments. Simulation results confirm our expectations and show that our system works pretty well in real room environments.

REFERENCES

[1] R. Aichner, S. Araki and S. Makino, "Time-domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," 12th IEEE Workshop on Neural Networks for signal processing, Sept. 2002, pp.445-454.
[2] 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.
[3] F. Asano, M. Goto, K. Itou and H. Asoh, "Real-time sound source localization and separation system and its application to automatic speech recognition," Proc. Eurospeech2001, pp.1013-1016.
[4] W. Baumann, D. Kolossa and R. Orglmeister, "Beamforming-based convolutive source separation," IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP'03), Vol. 5, 2003 pp: 357-360.
[5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE Trans. on Neural Networks, Vol. 10, No.3, pp.626-634, 1999.
[6] M. Z. Ikram D. R. Morgan, "A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation," ICASSP 2002, Vol. 1, pp.881-884.
[7] M. Knaak and D. Filbert, "Acoustical Semi-Blind Source Separation for Machine Monitoring," In ICA 2001, pp. 361-366.
[8] M. Knaak, S. Araki and S. Makino, "Geometrically constraint ICA for convolutive mixtures of sound," in Proc. IEEE ICASSP 2003, pp.725-729.
[9] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in Proc. IEEE ICASSP, 2000, pp.3140-3143.
[10] R. H. Lambert and A. J. Bell, "Blind separation of multiple speakers in a multipath environment," IEEE international conference on acoustic, speech and signal processing, ICASSP-97, Vol.1, Apr. 1997, PP. 423-426.
[11] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," Neurocomputing, Vol. 41, No. 1-2, pp. 1-24, Oct.2001.
[12] L. Parra and C. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," IEEE Trans. on Speech and Audio Processing, Vol. 10, No. 6, Sep. 2002, pp352-362.
[13] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain, " In Proc. ICASSP2004, vol. III, pp. 885-888, May 2004.
[14] 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. on speech and audio processing, Vol.12, No. 5, Sep. 2004.
[15] H. Saruwatari, T. Kawanura, K. Sawai, A. Kaminuma and M. Sakata, "Blind source separation based on fast-convergence algorithm using ICA and beamforming for real convolutive mixture," IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, (ICASSP '02), Vol.1, pp.921 -924.
[16] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., Vol. SP-34, No. 3, pp.276-280, March 1986.
[17] X. Sun and S.C. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," In Proc. ICA, San Diego, CA, Dec., 2001, pp.59-64.
[18] K.Torkkola, "Blind separation of convolved sources based on information maximization," IEEE Workshop on Neural Networks for Signal Processing, Japan,1996, pp. 423 –432.
[19] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag., Apr. 1988, pp.4-24.
[20] 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.