American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-06, pp-275-279
www.ajer.org

Research Paper Open Access

SCA based BASS: Using OMP

Shini P[1]., Ramya N[1]., Edet Bijoy K[2].,Muhammed Musfir N N[2].,
[1,2]*Department of Electronics & Communication Engineering,*
[1]*KMCT College of Engineering Calicut,* [2]*MES College of Engineering Kuttippuram,*

Abstract: - This paper deals with under-determined blind audio source separation, which is solved using sparse representations. The sparse component analysis (SCA) framework is a powerful method for achieving this. First, the mixing matrix is estimated in the discrete cosine transform (DCT) domain by a clustering algorithm. Then a dictionary is learned by an adaptive learning algorithm. Here, the Greedy Adaptive Dictionary (GAD) algorithm is utilized. Using the estimated mixing matrix and the learned dictionary, the sources are recovered adopting l_2-minimization technique called Orthogonal Matching Pursuit (OMP) as the sparse signal recovery method.

Index Terms: - *Blind audio source separation, Sparse component analysis, dictionary learning, sparsity.*

I. INTRODUCTION

Over the past two decades, blind source separation (BSS) has attracted a lot of attention in the signal processing community, owing to its wide range of potential applications, such as in telecommunications, biomedical engineering, and speech enhancement (Hyvarinen et al., 2001; Cichocki and Amari, 2003). BSS aims to estimate the unknown sources from their observations without or with little prior knowledge about the channels through which the sources propagate to the sensors. The instantaneous model of BSS can be described as:

$$X = A\,S \qquad (1)$$

where $A \in R^{MXN}$ is the unknown mixing matrix assumed to be full row rank, $X \in R^{MXT}$ is the observed data matrix whose row vector x_i is the ith sensor signal having T samples at discrete time instants t=1,…,T, and $S \in R^{NXT}$ is the unknown source matrix containing N source vectors. The objective of BSS is to estimate S from X, without knowing A.

Many algorithms have been successfully developed for blind source separation, especially for the exactly or over determined cases where the number of mixtures is no smaller than that of the sources. Independent component analysis (ICA) is a well-known family of BSS techniques based on the assumption that the source signals are statistically independent. However, ICA does not work in the underdetermined case, where the number of mixtures is smaller than that of the sources.

Underdetermined blind speech separation is an ill-posed inverse problem, due to the lack of sufficient observations, i.e. the number of unknown speech sources to be separated is greater than the number of observed mixtures. Several approaches have been developed to address this problem, such as the higher order statistics based method in (Comon, 1998), the sparse representations based technique in (Zibulevsky and Pearlmutter, 2001; Bofill and Zibulevsky, 2001). Good reviews on using sparse component analysis for source separation can be found in (Gribonval and Lesage, 2006; Sudhakar, 2011).

The key idea of sparse signal representation is to assume that the sources are sparse, or can be decomposed into the combination of a small number of signal components. By sparse, we mean that most values in the signal or its transformed coefficients are zero, except for a few nonzero values. These signal components are called atoms or code words, and the collection of all the atoms is referred to as a dictionary. Finding the sparsest representation (i.e. the non-zero coefficients) which best approximates the observation is often an NP-hard problem (Donoho, 2006).

In this work, the observed mixture is transformed by applying short-time DCT and the mixing matrix is estimated by clustering. Also here, sparse coding based on learned dictionary is used to solve the problem of

underdetermined blind speech separation. In particular, we propose a novel algorithm in which the BSS model is reformulated to a sparse signal recovery model. As a result, any of the state-of-the-art sparse signal recovery algorithms could be incorporated into this model to solve the underdetermined blind speech separation problem, with various separation performance and computational efficiency. This proposition was motivated by the failure of time domain algorithm called T-ABCD [1], in solving underdetermined BSS cases. It is an ICA framework along with k-means clustering [2], which found good results in determined cases.

II. PROPOSED METHOD

A. *Outline*

1. Apply short-time DCT to the mixture signal in X, for say, taking data frames of duration 25-35ms and an overlap of 10-15ms.
2. Estimate the mixing matrix, Â by K-means clustering of the normalized DCT coefficients.
3. A dictionary, Φ is learned on the transformed mixture signal by Greedy Adaptive Dictionary (GAD) algorithm.
4. Using Â and Φ, separate the sources by sparse signal recovery method, by reformulating the BSS problem into compressive sensing problem.
5. Reconstruction of separated sources by inverting the transform and the time domain signals are finally obtained.

The flow of the method is depicted in fig.1.

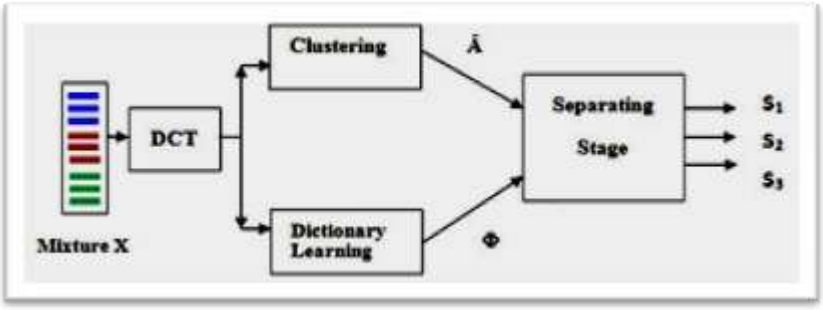

Fig .1 Flow of the proposed method.

B. *Steps in detail*

The transformed coefficients are undergone the three stage processing. It includes the mixing matrix estimation, dictionary learning and the source separation.

i. Mixing matrix estimation

The short-time DCT coefficients obtained in the first step are divided into *k* equal parts. Here *k* is equal to the number of sources and compute the mean values of each part as the initial centers. Run the K-means clustering algorithm to update iteratively the *k* centers until convergence and compute the column vectors of the estimated mixing matrix Â as the final centers.

ii. Adaptive dictionary learning

The dictionary atoms are obtained by using greedy adaptive dictionary (GAD) learning algorithm [3]. These obtained atoms can represent the features of the observed signal. GAD learns the dictionary atoms based on an iterative process using the sparsity index defined as follows:

$$\sigma_j = \frac{\|x_k\|_1}{\|x_k\|_2} \qquad (2)$$

where $\|.\|_1$ and $\|.\|_2$denote the l_1 and l_2- norm respectively and x_k is the column vector of the matrix containing the short-time DCT coefficients. The sparsity index measures the sparsity of a signal, where the smaller σ_j , the sparser the signal vector x_k. The GAD algorithm begins with the definition of a residual matrix R^d. This is first initialized to the transformed input matrix. The dictionary is then built by selecting the residual vector that has the lowest sparsity index. Then it is normalized and added to the dictionary. Finally, the new residual is computed for all the columns. The process is repeated until the number of obtained atoms reaches a pre-determined value.

iii. Separating sources by sparse signal recovery

In the separating stage, with the estimated mixing matrix Â, the underdetermined blind speech separation problem is formulated as a sparse signal recovery problem [4]. Equation (1) can be expanded as:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_M \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \dots & a_{MN} \end{pmatrix} \begin{pmatrix} S_1 \\ \vdots \\ S_N \end{pmatrix} \qquad (3)$$

where x_i (i=1,…,M) are the mixtures, s_j (j=1,…,N) are the sources, and a_{ij} is the ijth element of the mixing matrix A. Rewriting the above equation as follows,

$$\begin{pmatrix} x_1(1) \\ \vdots \\ x_1(T) \\ \vdots \\ \vdots \\ x_M(1) \\ \vdots \\ x_M(T) \end{pmatrix} = \begin{pmatrix} \Lambda_{11} & \dots & \Lambda_{1N} \\ \vdots & \ddots & \vdots \\ \Lambda_{M1} & \dots & \Lambda_{MN} \end{pmatrix} \begin{pmatrix} s_1(1) \\ \vdots \\ s_1(T) \\ \vdots \\ \vdots \\ s_N(1) \\ \vdots \\ s_N(T) \end{pmatrix} \quad (4)$$

where T is the length of the signal, $\Lambda_{ij} \in R^{TxT}$ is a diagonal matrix whose diagonal elements are all equal to a_{ij}. Let b= vec (X^T), f= vec (S^T), where vec is an operator stacking the column vectors of a matrix into a single vector. Equation (4) can be written in a compact form as:

$$b = M\, f \quad (5)$$

The above equation can be interpreted as a sparse signal recovery problem in a compressed sensing model, in which M is the measurement matrix and b is the compressed vector of samples in f. Therefore, a sparse representation in the transform domain can be employed for f:

$$f = \Phi\, y \quad (6)$$

where Φ is a transform dictionary and y contains the weighting coefficients in the Φ domain. Combining (5) and (6), we have

$$b = M\, \Phi y \quad (7)$$

In eq.(7) if y is sparse ,the signal f can be recovered from the measurement b using an optimization process. This indicates that source estimation in the underdetermined problem can be achieved by computing y in (7) using sparse signal recovery (i.e. sparse coding) methods.

Here the l_2 –minimization is adopted to find the sparse solution y. Specifically, OMP (Orthogonal Matching Pursuit) is used here. The orthogonal matching pursuit (OMP) (Pati et al.,1993) was developed to improve the MP (Matching Pursuit) by projecting the signal vector to the subspace spanned by the atoms selected as in MP via the same method. The basic idea of MP is to represent a signal as a weighted sum of atoms using Eq. (8) which involves finding the "best matching" projections of multidimensional data onto an overcomplete dictionary,

$$b = \sum_{i=1}^{k} y_i q_{\gamma_i} + r^{(k)} \quad (8)$$

where $r^{(k)}$ is a residual after k iterations, and q_{γ_i} is the atom of $M\Phi$ that has the largest inner product with the residual. At stage i, it identifies the dictionary atom that best correlates with the residual then subtract its contribution as follows,

$$r^{(i+1)} = r^{(i)} - y_i q_{y_i} \quad (9)$$

where $y_i = \langle r^{(i)}, q_{\gamma_i} \rangle$and $\langle , \rangle$ is an inner product operation. Then the process is repeated until the signal is satisfactorily decomposed. However, as opposed to MP, OMP maintains full backward orthogonality of the residual at each step when updating the coefficients:

$$b = \sum_{i=1}^{k} y_i q_{\gamma_i} + r^{(k)}, s.t. \langle r^{(k)}, q_{\gamma_i} \rangle = 0 \quad (10)$$

As proven in (Pati et al., 1993) the necessary number of iterations for OMP to converge is no greater than the number of atoms in the dictionary, while MP does not possess this property. The eq. (6) can then be written as:

$$\begin{pmatrix} s_1(1) \\ \vdots \\ s_1(T) \\ \vdots \\ s_M(1) \\ \vdots \\ s_M(T) \end{pmatrix} = \begin{pmatrix} D_1 & & \\ & \ddots & \\ & & D_M \end{pmatrix} \begin{pmatrix} y_1(1) \\ \vdots \\ y_1(T) \\ \vdots \\ y_M(1) \\ \vdots \\ y_M(T) \end{pmatrix} \quad (11)$$

where $D_1,...,D_M$ are identical dictionaries, $S_1,\ldots,S_M$ are the sources recovered and $y_1,\ldots,y_M$ are the sparse solutions. Finally, the estimates of separated sources are obtained by inverting the transform.

III. RESULTS AND DISCUSSION

The proposed algorithm was tested for various types of speech and music signals. For objective quality assessment three performance criteria defined in the BSSEVAL toolbox [4] was used to evaluate the estimated source signals. These criteria are the signal to distortion ratio (SDR), the source to interference ratio (SIR) and the source to artifacts ratio (SAR) [5], defined respectively as:

$$SDR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf} + e_{noise} + e_{artif}\|^2} \quad (12)$$

$$SIR = 10\log_{10}\frac{\|s_{target}\|^2}{\|e_{interf}\|^2} \quad (13)$$

$$SAR = 10\log_{10}\frac{\|s_{target} + e_{interf} + e_{noise}\|^2}{\|e_{artif}\|^2} \quad (14)$$

The experimental results are shown in the table 1.

TABLE 1 Separation performance measures

Sl.no.	Mixture	SIR (dB)	SAR (dB)	SDR (dB)
1.	I am_female_30s	0.0001	-79	0
	Poem _male	0.42	-44	-24.6
2.	Henry_theater_male	-3.16	-29	-44
	Poem _male	-165	-169	-156
3.	Music _signal_guitar	2.46	-28.6	-49.5
	Male _speech	0.00	-29	-26

The sources for the test are taken from [7]. The computational efficiency is improved when compared to STFT based and predefined dictionary based methods [6]. In general, separation performance highly depends on the mixing process. In this context, the accuracy of estimated mixing matrix is challenging.

The frame wise processing of data tremendously reduces the computation time whereas DCT provides good compression, so that less number of samples are undergone processing. The results are promising, even for this higher rate of compression. Also, the adaptive dictionary learns atoms with much faster rate compared to K-SVD [8].

IV. CONCLUSIONS

A multi-stage system for underdetermined blind speech separation using sparse coding with adaptive dictionary learning is presented. Numerical experiments have shown the competitive separation performance by the proposed method. The proposed method builds a new framework for underdetermined BSS, and offers great potential to accommodate the sparse signal recovery and adaptive dictionary learning algorithms to the source separation problems. This study has also shown the benefit of using learned dictionaries for underdetermined BSS, and the advantage of using the frame wise processing to improve the computational efficiency. Moreover, the framework of the proposed method provides a friendly structure to test the performance of other dictionary learning and signal recovery algorithms, specifically l_1 minimization techniques, in source separation applications in the future.

V. ACKNOWLEDGMENT

The authors acknowledge Ms. Jayasree T C, Head of the Department, Electronics & Communication Engineering, KMCT College of Engineering, Calicut and Dr.P. Janardhanan, Principal, KMCT College of Engineering, Calicut, for their support for fulfilling this work and their valuable comments for improving this paper.

REFERENCES

[1] Z. Koldovský and P. Tichavský, "Time-Domain Blind Separation of Audio Sources on the Basis of a Complete ICA Decomposition of an Observation Space", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 0, No.0, April 2010.

[2] K. Alsabti, S. Ranka, and V. Singh, "An Efficient K-Means Clustering Algorithm". *http://www.cise.ufl.edu/ ranka/*, 1997.

[3] Maria G. Jafari and Mark D. Plumbley, "Fast dictionary learning for sparse representations of speech signals", IEEE journal of selected topics in Signal Processing (special issue),2011.

[4] C. Févotte, R. Gribonval, and E. Vincent, BSS_EVAL toolbox user guide, IRISA, Rennes, France, Tech. Rep. 1706,2005, [online] Available: http://www.irisa.fr/metiss/bss_eval.htm

[5] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation", *IEEE Trans. Audio, Speech and Language Processing*, 14(4), pp 1462-1469, 2006.

[6] Tao Xu , Wenwu Wang , Wei Dai ,"Sparse coding with adaptive dictionary learning for underdetermined blind speech separation", Speech Communication 55 (2013) 432–450. http://www.telecom.tuc.gr/~nikos/BSS_Nikos.html

[7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representations," *IEEE Trans. on Signal Processing*, vol. 54, pp. 4311–4322, 2006.

[8] Maria G. Jafari, Samer A. Abdallah, Mark D. Plumbley and Mike E. Davies, "Sparse coding for convolutive blind audio source separation".

[9] Maria G. Jafari, Mark D. Plumbley and Mike E. Davies "Speech separation using an adaptive sparse dictionary algorithm", IEEE HSCMA 2008.

[10] Andrew Nesbit, Maria G. Jafari, Emmanuel Vincent and Mark D. Plumbley, "Audio Source Separation using Sparse Representations", Machine Audition: Principles, Algorithms and Systems IGI Global (Ed.) (2010) pp.246—265.