INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

# A COMPRESSIVE SENSING APPROACH TO SPEECH SEGREGATION

Swapnil Mohan Mahajan[1*], Chetankumar Bhogayta[2]
[1*2]Student, M.Tech Communication Engineering, Vellore Institute of Technology, Chennai
Correspondence Author: Swapnil.mohan2013@vit.ac.in

**Keywords:** Compressive Sensing; Sparsity; GPSR; K-means; L1-magic.

## Abstract

The problem of underdetermined blind source separation is usually addressed under the framework of sparse signal representation. This paper represents Compressive Sensing technique used for speech segregation that contains two stages. In the first stage we exploit a modified K-means method to estimate the unknown mixing matrix. The second stage is to separate the sources from the mixed signals using the estimated mixing matrix. In the second stage a two-layer sparsity model is used which assumes that the low frequency components of speech signals are sparse on K-SVD dictionary and the high frequency components are sparse on discrete cosine transformation (DCT) dictionary. In this way, we reconstruct the signals using L1-magic and GPSR algorithm**.**

## Introduction

The task of BSS [1] is to recover the sources using observable sources. Here is the noise free mixing model of BSS is described as follows:

$$\mathrm{x(t)} = \mathrm{A}\, s(t) \qquad (1)$$

Where, A ∈ $R^{M \times N}$ is the mixing matrix which is unknown, S(t) ∈ $R^N$ is original unknown source vector and X(t) ∈ $R^M$ is the observed mixed data vector at discrete time instants. In this paper, we are going to discuss about undetermined case [1], [2] of BSS i.e., M < N. For the simplicity I have chosen M = 2 and N = 3.

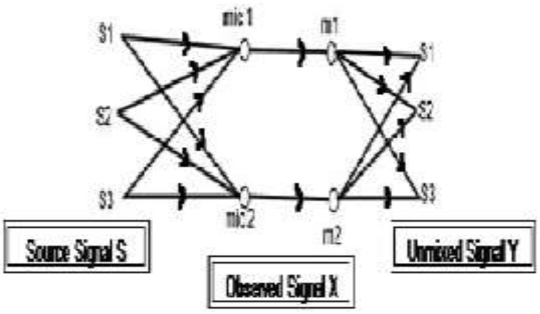

***Fig 1 A diagrammatic representation of a mixing and an unmixing system***

From fig.1, it is clear that the undetermined problem is encountered where the number of the sources is greater than that of the mixtures. For this an effective method for this problem is to use the so-called sparse signal representation [3], assuming that the sources are sparse.

**Preliminary work**

To the simplest of my data, most of the proposed methods rely on the sparsity of signals in some domain, like the time-frequency (TF) domain. Some authors conjointly tried to use trained dictionaries to replace discrete cosine transformation (DCT) or fast Fourier transformation (FFT) dictionary which is fixed, and showed that the trained dictionaries perform better than the fixed dictionaries.

Based on such a representation, a two-stage approach is usually used to recover the source signals. First, the unknown mixing matrix is estimated from the audio mixture data by assuming a mixing matrix randomly. Second, for the underdetermined problem we propose a new method for the source recovery in the second stage based on the emerging technique of compressed sensing (CS) [1], [2], [3]. The CS, that has attracted growing interests in signal process, is an efficient technique for data acquisition and reconstruction. It can randomly sample signals under Nyquist rate and then reconstruct the signal with a high probability. It provides potentially a powerful framework for computing a sparse representation of signals. It can randomly sample signals under Nyquist rate and then reconstruct the signal with a high probability. It provides potentially a strong framework for computing a sparse representation of signals. In this work, we analyse the similarity between the fundamental models for CS and BSS, and so develop an algorithm for source recovery based on their relations.

## Compressive sensing & sparsity

### Compressive sensing

Compressive Sensing [3] is one of the unique techniques of acquiring signals. CS provides for each sampling moreover as compression, in conjunction with coding of the supply data, at the same time. Also, signal reconstruction is also helps to recover our signal. These all advantages of CS are very important for the communication purpose. For Compressive Sensing, signals should be sparse.

The fundamental mathematical formulation of CS is discussed here to build a CS framework for our demand. Let us assume that a 1-D signal having N samples is measured in transform domain given by a linear mapping

$$y = Ax \qquad (2)$$

Where, $A$ is consist of basis functions of N x N matrix such that

$$x = A^{-1}y \qquad (3)$$

In this case during measurement if only K samples of $y$ is measured and even if K is much lesser than N, CS frame work allows us to reconstruct $x$ from the incomplete information of $y$ exploiting the sparse nature of $x$. It is shown that in ref, if $x$ is S sparse and if the following rule (4) is also obeyed,

$$K \geq C.S.logN \qquad (4)$$

Where, C is a constant.

### Sparsity

The selection of sparsifying operator B is that the question within the case of compressive sensing of speech and therefore the previous works are to define this B and an approach to resolve equation (6). It is shown that there is no guarantee that the operations like DFT, Wavelet decomposition etc sparsify the speech signal better than the classic speech analysis models like LPC analysis.

Here, we are going to discuss about sparsity [4] of speech signal. So, sparsity means signals that are largely inhabited with zeros and have a small variety of non-zero elements known as sparse signals. A sparse representation of signal is the one where small number of coefficients contains large proportion of energy.

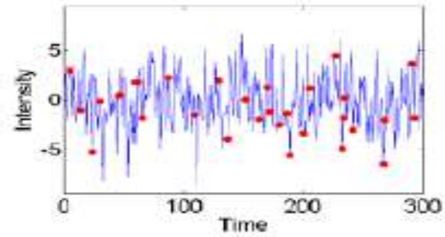

***Fig. 2 Observed Signal***

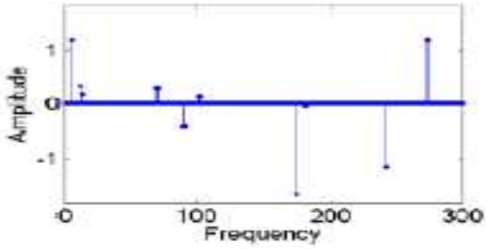

***Fig. 3 Sparse Signal***

Fig. 2 represents the original samples signal which is highly populated with non-zero samples whereas fig. 3 represents its Fourier transform which contain large amount of zero samples. This is known as representation of sparse signal [4].

## Estimating the mixing matrix by k-means algorithm

K-means clustering [1] is a method of vector quantization originally from signal processing. The K-means algorithm is an algorithm to cluster *n* objects based on attributes into *k* partitions, where $K < n$. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid.

Previously the mixing matrix is unknown. In this stage, we are going to estimate the mixing matrix by using singular value decomposition (SVD) [1] and K-means clustering [1] algorithm. In TF domain,

$$X = AS \qquad (5)$$

Where, X and S are the TF coefficient of x(t) and s(t) respectively. At every TF point (w, t), we have

$$\begin{bmatrix} X_1(w,t) \\ X_2(w,t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{11} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} S_1(w,t) \\ S_2(w,t) \\ S_3(w,t) \end{bmatrix} \qquad (6)$$

As a signal is sparse, at a large proportion only one source is active so that $X_1(w,t)/X_2(w,t)$ is approximately equal to $a_{1i}/a_{2i}$, $i$ = 1, 2, 3. Hence a scatter plot of $X_1$ vs. $X_2$ would cluster into three distinct lines such that $ith$ source corresponds to the line with gradient $a_{1i}/a_{2i}$, $i$ = 1, 2, 3. Then we can use the K-means algorithm to obtain three clusters and estimate every column of matrix with an amplitude uncertainty.

But, the assumed sparsity is hard to specify for all points. Here, we exploit SVD instead of histogram. We define the covariance matrix of TF mixture vectors:

$$R_X = E[XX^T] = AR_S A^T \qquad (7)$$

Where, $X = \begin{bmatrix} X_1(w,t) \\ X_2(w,t) \end{bmatrix}$ and we exploit SVD to $R_X$ :

$$R_X = a_i s_i^2 a_i^T \qquad (8)$$

Where, $U = [u_1, u_2]$ is the eigenvector matrix and $\Sigma = diag\{\sigma_1^2, \sigma_2^2\}, \sigma_1^2 \geq \sigma_2^2$ is the eigenvalue matrix. Here we estimate $R_X$ by frequency average in a ST window, i.e., $E\{XX^H\} \approx (1/l)\Sigma_w X(w,t)X^H(w,t)$, where $l$ is the length of ST window. If only the $ith$ source is active in some ST window, then

$$s_i^2 \neq 0, s_{i'}^2 = 0, (i' \neq i), \qquad (9)$$

$$R_x = a_i s_i^2 a_i^H \qquad (10)$$

Where, $s_i^2$ is the power of $ith$ source. Under the assumption, the rank of $R_X$ is 1 and $\sigma_1 > 0, \sigma_2 = 0$. Using this, (7) can be simplified as:

$$R_X = \sigma_1^2 u_1 u_1^H \qquad (11)$$

Comparing (10) and (11), we can see that, $u_1$ is the estimate of $a_i$. We find that every ST window is corresponds to eigenvector $u_1$ and all of the $u_1$ cluster into three different vectors corresponding to the columns of mixing matrix.

But, sparsity assumption is not always satisfied. If we use all the three clusters of $u_1$, an inaccurate estimate will be got. There is a condition, whose corresponding ST window is not sparse, i.e., $1-(\sigma_2/\sigma_1) > ebs$ where $ebs$ denotes a real number close to 1. Therefore, we only use the reliable cluster samples and we can get a better estimate. It should be stated that, the problem of scaling and permutation uncertainty exists in the estimated matrix.

## Methodology

### GPSR (Gradient projection of sparse reconstruction)

GPSR [5] algorithm is one of the best techniques to reconstruct the sparse signal. Many problems in signal processing and statistical inference involve finding sparse solution to under-determined linear system of equation. Basis Pursuit, the Least Absolute Shrinkage and Selection Operator (LASSO), wavelet based deconvolution, and Compressed Sensing is a few well-known examples of this approach.

It can be reconstructed with an overwhelming probability by minimizing $l_1$ norm of $x$ and can be modeled as a convex unconstrained optimization problem given by

$$\underbrace{min}_{x} \frac{1}{2}\|y - Ax\|_2^2 + \tau\|x\|_1 \qquad (5)$$

Where $x \in \mathbf{R}^n$, $y \in \mathbf{R}^k$, A is an $k \times n$ matrix, τ is a nonnegative parameter. GPSR [5] is able to solve a sequence of problems efficiently for a sequence of values of τ. Once a solution has been obtained for a particular τ, it can be used as a "warm-start" for a nearby value. Solutions can therefore be computed for a range of τ values for a small multiple of the cost of solving for a single τ value from a "cold start."

This assumes x is sparse in time domain. If x is sparse in a transform domain $x = Bs$ where as $B$ is also a N x N basis matrix and s is sparse then the objective function becomes

$$\underbrace{min}_{x} \frac{1}{2}\|y - ABs\|_2^2 + \tau\|s\|_1 \qquad (6)$$

Here A is called sensing matrix and B is called sparsifying matrix and it is proved that A and B to be mutually non coherent for better reconstruction of x.

There is alternative interpretation that includes operator A and B that works on x to produce the transformed output. Large scale implementation of CS algorithm needs implementation of these operators which are used to iteratively solve the optimization problem in (6). This helps the implementation to avoid representation of complex and large A and B matrices and also helps to use the prevailing quick blocks to see the transforms.

**Results**

We are able to separate original signals from mixed signal. To implement this we used three speech signals such as Bat, Boot, But respectively.

Original signals:

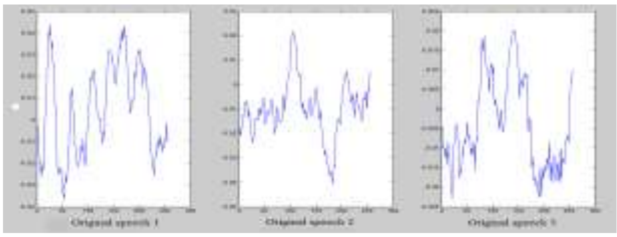

*Fig. 4 Plot of three originals speech signals*

Mixed signals:

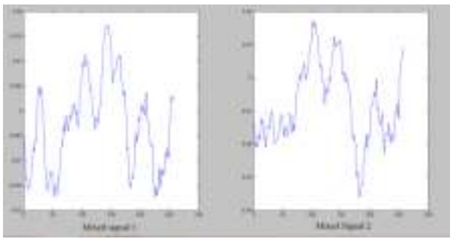

*Fig. 5 Plot of mixed signals*

Separated signals:-

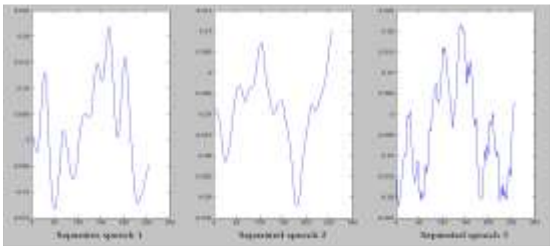

*Fig. 6 Plot of three separated speech signals*

**L1-magic**

L1-Magic [6] is employed for recovery of sparse signals. The central results state that a sparse vector $x_0$ $R^N$ may be recovered from a small number of linear measurements y = $Ax_0$ $R^K$ where, K<<N by solving a convex program.

For maximum computational efficiency, the solvers for every of the seven issues are implemented one by one. All of them have a similar basic structure, however, with the procedure bottleneck being the calculation of the Newton step. The code may be used in

either “small scale” mode, where the system is constructed explicitly and solved exactly, or in “large scale” mode, where an iterative matrix-free algorithm such as conjugate gradients (CG) [6] is used to approximately solve the system.
For our experiment, we used ”Min- ℓ1 with equality constraints” issue [6]. To illustrate this, the ℓ1-Magic package includes m-files for solving specific instances.

**Results**

Original Signals:

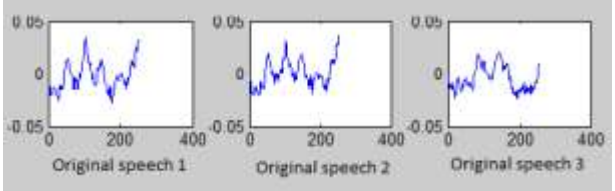

***Fig. 7 Plot of three originals speech signals***

Mixed Signals:

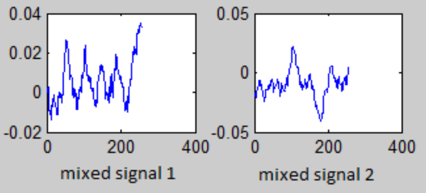

***Fig. 8 Plot of mixed signals***

Separated Signals:

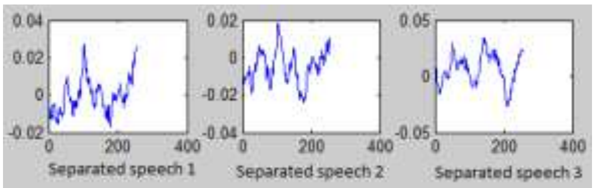

***Fig. 9 Plot of three separated speech signals***

## Experimental results

In our simulations, we generate two mixture signals by mixing three audio sources. The sources include a signal containing “bat”, “boot”, “but” speech respectively.
Previously, we assumed the unknown mixing matrix by considering matrix such as,

$$A = \begin{pmatrix} 0.9 & 0.6 & 0.2 \\ 0.1 & 0.3 & 0.4 \end{pmatrix}$$

Where the cluster points of each column matrix is [0.9 2 0.5]. But after applying K-SVD and K-means algorithm, we

$$A = \begin{pmatrix} 0.66 & 0.88 & 0.49 \\ 0.72 & 0.44 & 0.86 \end{pmatrix}$$

got the mixing matrix using three cluster values and also find that the both matrix’s cluster points are close to each other. Now the known mixing matrix is
Where the cluster points of each column matrix is [0.91 2 0.57]. The scatter plot of three clusters is shown in following fig 10.

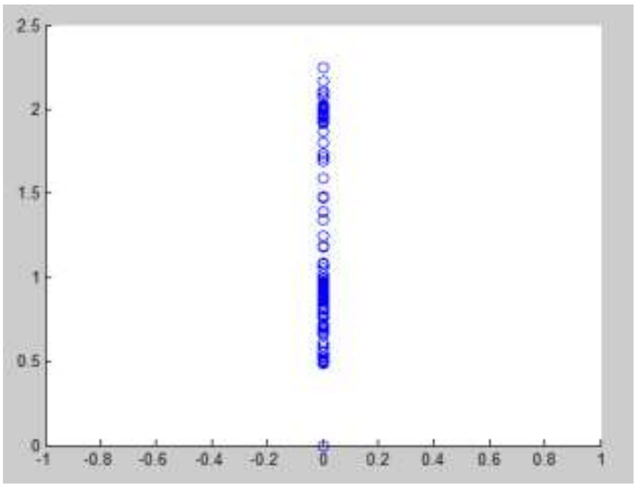

***Fig. 10 Scatter plot of three cluster points of estimated mixing matrix***

With the help of this known mixing matrix we can go further for separation process. Both the methods are able to get the signals back from mixed signal. By comparing both algorithm, GPSR is better than L1-Magic at 'tau' value 0.015.

## Conclusion

In this paper, we proposed a CS approach to undetermined BSS which contains two stages. Experiment showed that after applying K-SVD and K-means algorithm we get the known mixing matrix which is used to recover the original source signals from mixtures. Furthermore, GPSR performs better than L1-Magic.

## Acknowledgment

This project is result of the hard work put in by us. This project has been a great learning experience for us and its timely completion would not have been possible without the help of our school faculties and facilities provided by VIT management.

## References

1. Guangzhao Bao, Zhongfu Ye, Xu Xu, Yingyue Zhou, "A Compressed Sensing Approach to Blind Separation of Speech Mixture Based on a Two-Layer Sparsity Model", IEEE transactions on audio, speech and language processing, vol.21, No 5, May 2013.
2. Tao Xu and Wenwu Wang, "A Compressed sensing approach for undetermined blind source separation with sparse representation" Centre for Vision, Speech and Signal Processing University of Surrey, Guildford, United Kingdom, 2009 IEEE.
3. T.V. Sreenivas and W. Bastiaan Kleijn " Compressive Sensing for Sparsely excited speech signal" Department of ECE, Indian Institute of Science, Bangalore-12, India , ACCESS Linnaeus Center, Electrical Engineering KTH - Royal Institute of Technology, Stockholm, © 2009 IEEE.
4. Siddhi Desai, Prof. Naitik Nakrani, "Compressive Sensing in Speech Processing: A Survey Based on Sparsity and Sensing Matrix," International Journal of Emerging Technology and Advanced Engineering, :2008 Certified Journal, Volume 3, Issue 12, December 2013).
5. M´ario A. T. Figueiredo, Robert D. Nowak, Stephen J. Wright, "Gradient rojection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," Submitted for publication; 2007.
6. Emmanuel Cand`es and Justin Romberg, Caltech, "ℓ1-magic : Recovery of Sparse Signals via Convex Programming," October 2005.http://users.ece.gatech.edu/~justin/l1magic/