Compressive Sensing and its Application to Speech Signal Processing

Pooja C. Nahar, Dr. Mahesh T. Kolte

Department of Electronic & Telecommunication MIT College of Engineering
University of Pune Pune,India
poojanah@gmail.com

Department of Electronic & Telecommunication MIT College of Engineering
University of Pune Pune,India
mahesh.kolte@mitcoe.edu.in

***Abstract*—Compressed sensing or compressive sensing or Compressive sampling or CS is a revolutionary technique that strongly relies on sparsity of signal. CS has been applied in areas like Radar systems, sonar systems, video signal, and image processing etc. This paper will discuss the potential implementation of CS technology on speech signal and the compression rate that can be achieved.**

***Index Terms*—Compressive sensing, speech signal processing, Compressive sampling, applications of CS, data acquisition.**

Introduction

Conventionally, the signal of interest is sampled at nyquist rate which states that, the minimum frequency of sampling a particular signal is twice that of the maximum frequency content present in that signal. Sampling at this frequency leads to production of large amount of samples, out of which, most are eliminated during the compression stage. This leads to unnecessary hardware and software load. Compressed sensing, as the name suggest, samples the signal in a compressed format i.e. it uses very less number of distinct samples of the target signal and is then recovered by using various recovery algorithms. As a result, less number of samples are handled, which leads to reduction in power consumption as well as a reduced load on hardware as well as software

The basic pre-requisite for a signal to be compressively sensed is the sparsity. Sparsity means, more number of components in a signal are equal to or nearer to zero and less number of components are distinct. Also, the other pre-requisite is incoherence, which pertains to the sensing modality.

Most of the natural signals present around us are sparse in nature. If, we consider speech signal, it is sparse in nature, and hence, we can apply CS to speech signal. Also, the noise present in speech signal is non-sparse, hence we can remove the noise present in speech signal using CS as CS will not be able to process the non-sparse component i.e. noise.

In this paper, a speech signal was recorded in matlab, converted into sparse signal and then CS was applied on to it in order to get the signal to be transmitted. Then, at the receiver end, the signal was recovered using l1 minimization algorithm and then inverse transformed in order to recover the synthesized speech signal.

This paper is organized as follows. Section 2 provides an overview of CS, section 3 gives an idea on implementation of CS on speech signal, section 4 describes results and section 5 summarizes the findings and future work

Background

A. *Overview of Compressed Sensing*

The theory of CS was developed by Candes [1] and Donoho [2] in 2004. It involves taking random projections of the signal and recovering it from a small number of measurements using optimization techniques. In a traditional sampling theorem, the signal is sampled using Nyquist rate, whereas with the help of compressive sensing the signal is sampled below the Nyquist rate.

This is possible because the signal is transformed into a domain in which it has a sparse representation. Then the signal is reconstructed from the samples using one of the different optimization techniques available

For understanding the concept of compressed sensing, we will go through following set of definitions and formulae:

i. **Sparse Signal:** A signal is called sparse in nature if it has only a few large in magnitude components and a greater number of components which are close to zero.

ii. **Compressible Signal:** A signal which has a sparse representation is compressible.

iii. $s = \Psi x$ where, s = Signal to be acquired
Ψ = Sparsifying matrix
x = Real valued Column vector

iv. $y = \Phi s = \Phi\Psi x = Ax = A_k . x_k$ where, y = Compressed Samples
Φ = Sensing Matrix

v. The Solution to above equations is:

$$X_k = (A_k^T A_k)^{-1} A_k^T y$$

vi. Above is an underdetermined problem i.e. projection of an n-dimensional vector into an M dimensional space i.e. Number of equations < Number of Unknowns

vii. To solve this kind of problems, we use the concept of Norms. Norms are nothing but, they assign strictly positive length to vectors in a vector space. Norms are of following types:

a. L_0 Norm: It is used to count the number of non-zero components present in a vector
b. L_1 Norm: It is given by the following equation:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i|$$

c. L_2 Norm: It is given by following equation:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{N} |x_i|^2 \right)^{\frac{1}{2}}$$

viii. **Designing a Sensing Matrix:** Following conditions need to be strictly satisfied while designing a sensing matrix so that, the signal is recovered faithfully:

a. Universal Incoherence condition: It means that, the value of cross correlation between two column vectors of a sensing matrix must be minimum.
b. Data Independence: The sensing matrix need not be dependent upon the prior knowledge of data.
c. Robustness: Randomly projected coefficients are robust to packet loss.

ix. **Incoherence condition:** The sensing matrix should be as different from the sparsifying basis. Time and frequency basis are maximally incoherent. Following equation signifies the incoherence condition:

$$\mu < 1/(2K-1)$$

RECOVERY ALGORITHMS

There are basic two types of approaches as follows:

1. Greedy Type – Orthogonal Matching Pursuit

OMP is a greedy-type algorithm. It performs a greedy search mechanism for each iteration to find the sparse set. The solution will be wrong if it fails to reconstruct the sparse set exactly. It is mostly used when the number of common components is more.

2. Gradient Type – Primal Dual Interior Point

This algorithm is nothing but making some changes into the L_1 norm and is mainly used when the innovation components are more. In this paper, we will be using primal dual interior point algorithm.

APPLICATIONS OF CS TO SPEECH SIGNAL

The following figure 1 explains the basic block diagram of compressed sensing applied to speech signal.

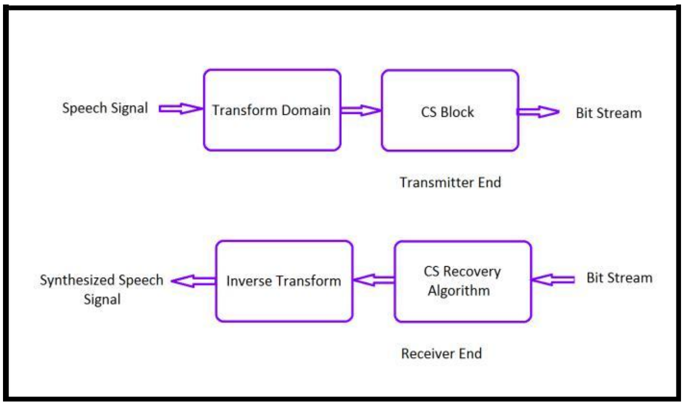

Fig. 1: Block diagram of CS applied to speech signal

First, a speech signal of 2seconds is recorded in matlab. Then, the recorded speech signal is first transformed into a domain in which it is sparse. Here, the transform domain used is Discrete Cosine Transform (DCT). Next step is to design a window function that will be used to multiply all the components in that window by zero. This will make the transformed signal sparser. The window used is -0.06 to +0.04. The resulting signal is ready to be sensed compressively by making random linear projections onto the signal. This is done using a random matrix of size K by N. here, K is the sparse number and, N is the total number of samples in the transformed and windowed function.

By using random matrix, we make random linear projections on the sparse signal in order to take very few components of the sensed signal. Thus, the signal to be recovered becomes robust for any errors. Hence we now have the compressively sensed signal y(also called as observation vector). This observation vector is recovered using L1 minimization algorithm at the receiver and inverse transformed in order to recover the original speech signal.

RESULTS

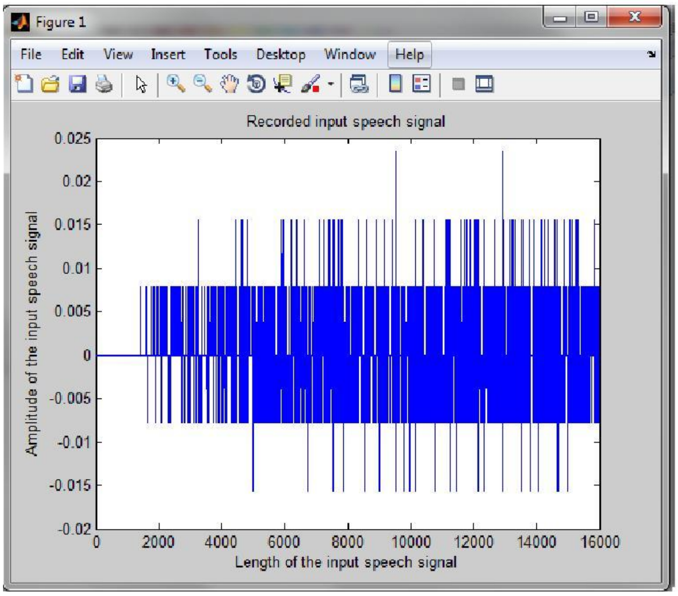

Fig 2: Recorded speech signal

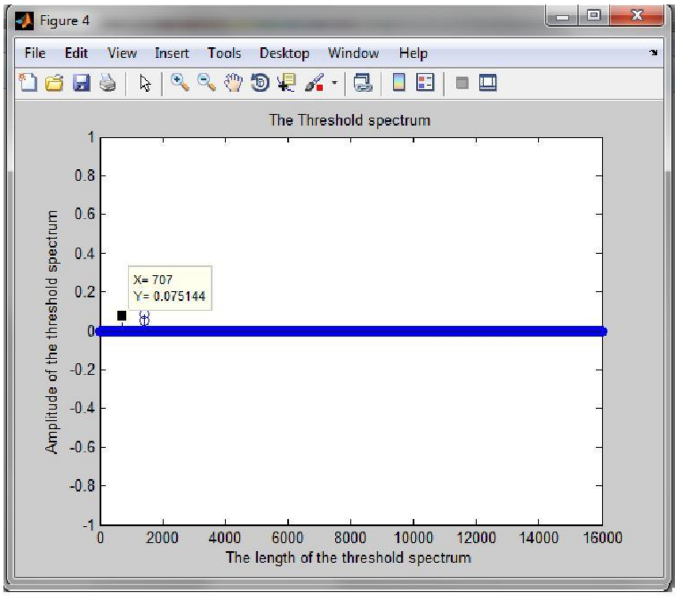

Fig 4: Threshold Spectrum

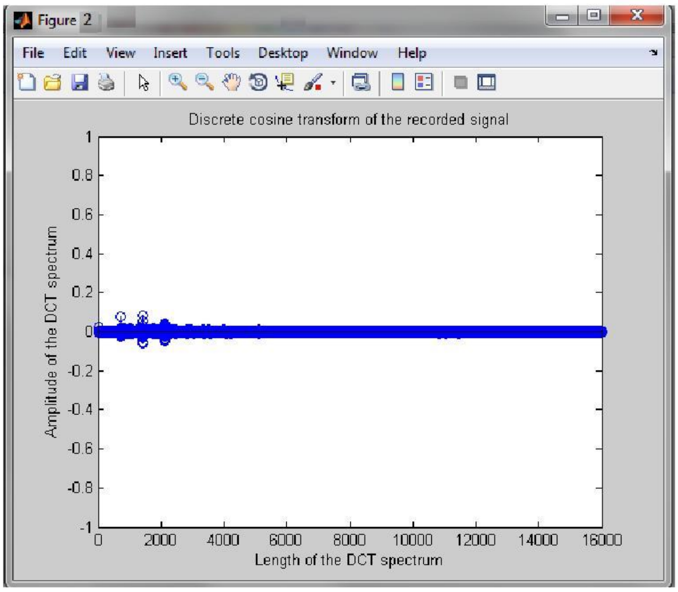

Fig 3: DCT of recorded speech signal

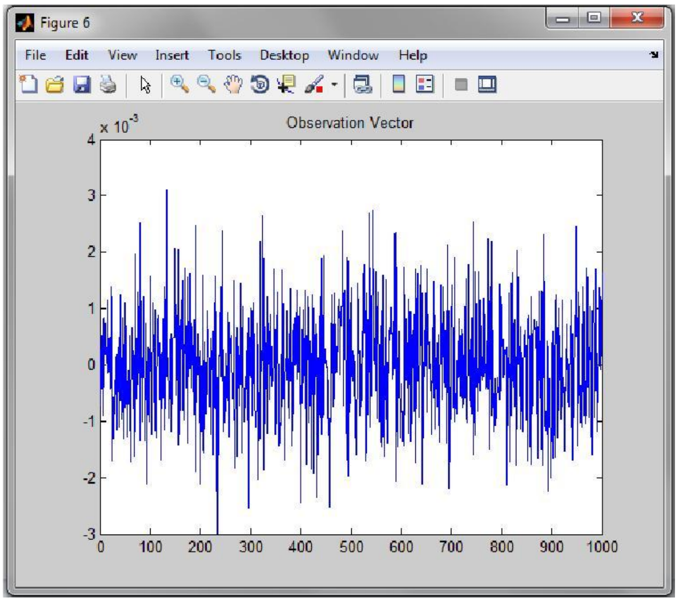

Fig: 5: Observation vector

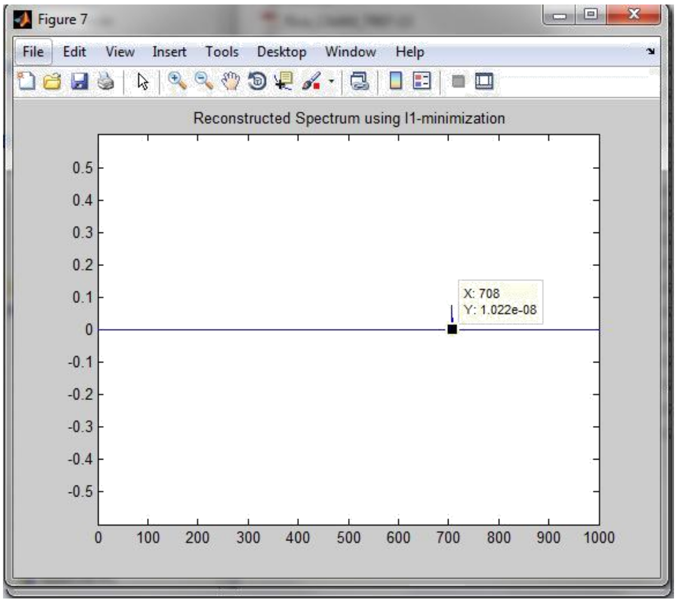

Fig: 6: Reconstructed spectrum using L1 minimization

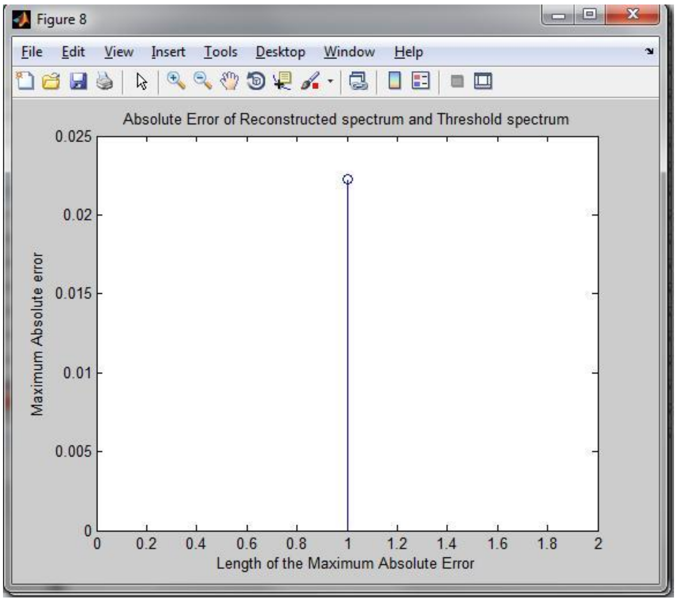

Fig 7: Maximum absolute error in reconstructed spectrum

From the above figures, following points are clear:

a. Figure 2 is a recorded speech signal of duration 2seconds and contains the utterance of letter „A‟.
b. Figure 3 is nothing but the Discrete cosine transform of the recorded speech signal.
c. Figure 4 is the thresholded spectrum. The threshold window used is from -0.06 to +0.04. Any components in this window are made equal to zero. This is done in order to make the signal sparser.
d. Figure 5 is nothing but the observation vector. It is the compressively sensed signal obtained after applying the random linear projections onto the thresholded spectrum.
e. Figure 6 is the reconstructed spectrum using L1 minimization algorithm. We can see from the figure that, it has a peak of value 0.075 at x = 707 which is same as the value present in the thresholded spectrum. This means that, we have accurately recovered the spectrum using L1 minimization algorithm.
f. Figure 7 shows the maximum absolute error that is present between the thresholded spectrum and the recovered spectrum using L1 minimization algorithm. The maximum absolute error is 0.023.

The above results are by using values of K (sparse number) as 1000 out of total number of samples present i.e. 16000. Hence the degree of compression is huge. We can vary the sparse number and observe the results. Following table gives the variation in output error on varying the sparse number K:

N	K	Maximum absolute Error	Compression (%)
16000	500	0.70	96.875
16000	1000	0.14	93.75
16000	5000	0.06625	68.75
16000	10000	0.04631	37.5

CONCLUSION

CS can prove to be a revolutionary technique for signal acquisition and recovery. The key advantages are:

- Fast acquisition of data with fewer samples
- Decreased computational complexity
- Lower transmission power and higher compression rates
- Small traffic volume
- Small time delay
- Sampling matrix need not be adaptive to signal
- The desired resolution for recovering the compressively sensed signal can be achieved by manipulating the sparse number K.

FUTURE WORK

Recover signal using different recovery algorithms such as OMP and Basis pursuit algorithm etc. in order to recover original speech signal.

Use different transform domains and check the output.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr.2006.

[3] "An Introduction To Compressive Sampling" by Emmanuel J. Candès and Michael B. Wakin IEEE Signal processing magazine March 2008

[4] Siow Yong Low a, Duc Son Pham b, Svetha Venkatesh c "Compressive speech enhancement" Science Direct\ Speech Communication vol 55 pp. 757–768, Feb 2013

[5] Paliwal, K., Wojcicki, K., Schwerin, B., 2010. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. Speech Communication 52 (5), 450–475.

[6] ITU, 2001. Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU Recommendation, 862.

[7] O"Shaughnessy, D., 2000. Speech Communications: Human and Machine. IEEE Press, NJ, USA.

[8] "Robust Speech Recognition Using a Cepstral Minimum-Mean-Square-Error Motivated Noise Suppressor" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 16, NO. 5, JULY 2008

[9] "Voice Quality Solutions for Wireless Networks" White Paper, March 2012

[10] "SPEECH PROCESSING A Dynamic and Optimization-Oriented Approach" Published by Marcel Dekker Inc. New York, NY, U.S.A.

[11] "Audio Signal Processing and Coding" by Andreas Spanias, Ted Painter and Venkatrman Atti

[12] "Robust Speech Recognition for Adverse Environments" by Chung-Hsien Wu and Chao-Hong Li

[13] "Scalable Video Coding with Compressive Sensing for Wireless Videocast" by Siyuan Xiang and Lin Cai

[14] "Intelligent Sensor Networks: Across Sensing, Signal Processing, and Machine Learning" by Jae-Gun Choi, Sang-Jun Park, and Heung-No Lee