International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.4 (2014), pp.371-380
http://dx.doi.org/10.14257/ijsip.2014.7.4.35

Improved Performance of Compressive Sensing for Speech Signal with Orthogonal Symmetric Toeplitz Matrix

Desai Siddhi[1] and Nakrani Naitik[2]

[1]PG Student, Electronics & Communication Department, CGPIT, Bardoli
UTU, Gujarat, INDIA
[2]Assistant Professor, Electronics & Communication Department, CGPIT, Bardoli
UTU, Gujarat, INDIA
[1]siddhi_desai@ymail.com, [2]naitiknakrani@gmail.com

Abstract

In Compressed Sensing (CS) framework, reconstruction of a signal relies on the knowledge of the sparse basis & measurement matrix used for sensing. Most of the studies so far focus on the application of CS in fields of images, radar, astronomy and Speech. This paper introduce new sensing matrix called orthogonal Symmetric Toeplitz Matrix (OSTM) generated with Binary, Ternary and PN sequence and shows detailed comparison of them with DCT Basis applied on 8 KHz sampled speech signal. Also it shows improved results of OSTM compared to Random, Bernoulli, Hadamard and Fourier Matrices. Performance of sensing matrices has been compared with Mean square error, Signal to noise ratio and Perceptual Evaluation of Speech Quality (PESQ) parameters.

Keywords: *Discrete Cosine Transform (DCT), Orthogonal Symmetric Toeplitz Matrix (OSTM)*

1. Introduction

Compressive sensing or C.S. is a very simple, efficient, non adaptive and parallelizable compressed data acquisition protocol that provides both sampling and compression along with encryption of source information simultaneously. The theory of compressive sensing was developed by Candes, *et al.*, and Donoho in 2004 [8]. This method is different from traditional method as it sampled the signal below the Nyquist rate and it permits to exploit the sparse property at the signal acquisition stage of compression.

In compressive sensing, the signal is first transformed into a sparse domain and then the signal is reconstructed using numerical optimization technique using small number of linear measurements. Implementation of Compressive sensing Theory in specific application reduced sampling rates, or reduced use of Analog to Digital converter resources. Compressive sensing is a new paradigm of acquiring signals, fundamentally different from uniform rate digitization followed by compression, often used for transmission or storage [1-3].

Compressive sensing can be used in many applications, especially speech processing. It has been used in noise reduction, speech denoising and speech coding [6]. However, as it is still a new technology, not much research has been done on the use of CS for speech signal.Speech is a natural way of communication between two persons but its processing is difficult because Even if we utter a same word we can’t produce the same signal ever in our entire life. Thus major challenges to apply CS in speech processing starts with, designing of efficient measurement matrices & finding a good sparse basis.

ISSN: 2005-4254 IJSIP
Copyright ⓒ 2014 SERSC

Compressed sensing requires the measurement matrices be incoherent with the sparse basis. The minimum the coherence is, the less the required number of measurements is, the more information in the sensed signal will be contained, and the higher the probability of reconstruction is, therefore, an efficient method of satisfying optimal incoherence could be constructed for the design of measurement matrix. One popular family of sensing matrices are random Gaussian/Bernoulli matrices. However, it takes huge memory buffering for storage and its high computational complexity due to unstructured nature restrict its applications for speech processing. As a result, researchers have proposed some methods to construct deterministic matrices like Fourier, Wavelets, *etc.*, [11]. The objective of this paper is to introduce a new deterministic sensing matrix named Orthogonal Symmetric Toeplitz Matrix (OSTM) generated by Binary, Ternary and PN sequence. Comparative analysis of various sensing matrices has been done in this work.

This paper is organized as follows. This section gives an introduction about compressive sensing. In section II, a review about compressive sensing theory is presented. Different sensing matrices that are used for analysis are presented in this section. Procedure for generating Orthogonal Symmetric Toeplitz matrix is explained in section II. In section III analysis of Compressive sensing for speech compression application is done with different sparsity basis. The conclusion is given in section VI.

2. Compressive Sensing Basics

The basic principle of Compressive Sensing is shown in Figure 1. It consist two main parts: transmitter and receiver. Transmitter side input signal x is given with N samples. First x has to be converted into some domain in which x has sparse representation. For example, DCT, DFT *etc.*, after this conversion signal x is transformed into K – sparse signal. Where K is largest coefficients obtained using thresolding. These K largest coefficients contain most of the information about signal. Then it is multiplied with sensing matrix ϕ and result will give M – length measurement matrix.

At the receiver side, different optimization techniques are used for reconstruction of original signal. First multiplication of signal with sensing matrix is computed which gives N samples from M measurements. Then convex optimization techniques are used to recover K-sparse signal. Once again inverse sparsity is applied to obtain original signal [13].

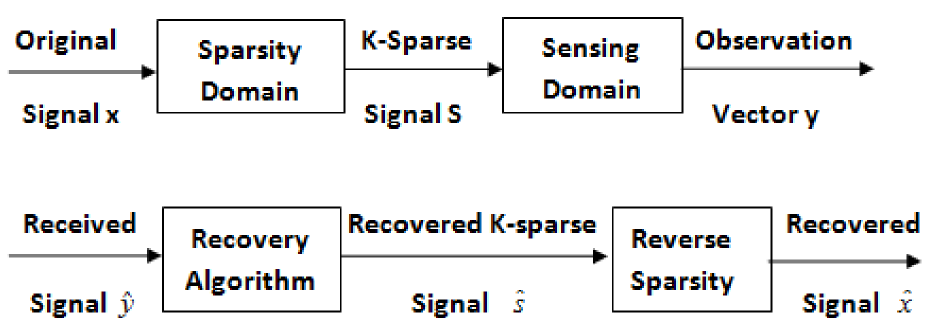

Figure 1. Block Diagram

In short working of compressive sensing theory is mathematically expressed by following manner:

 Copyright ⓒ 2014 SERSC

Let $x \in R^N$ be the speech signal and let $\psi = [\psi_1, \psi_2, \ldots \psi_N]$ be the basis vectors spanning. The speech signal is said to be sparse if,

$$x = \psi \cdot s = \sum_{i=1}^{k} s_{ni} \cdot \Psi_{ni} \ [n_1, n_2, \ldots n_k] \subset [1, \ldots N] \tag{1}$$

Where, s_{ni} are scalar coefficients and $K << N$, *i.e.*, s_n or simply s is the sparse vector with only K non-zero elements. Based on CS theory, perform sampling of x through projections onto random bases and reconstruct the speech signal at the receiver with full knowledge of the random bases.

In other words, the sampling (sensing) measurements can be defined as:

$$y_m = \sum_{i=1}^{N} \phi_m(i) x(i), \quad 1 \le m \le M < N \tag{2}$$

Or $y = \phi \cdot x$, where $\phi = M \times N$ is measurement matrix. The ϕ is made up of orthonormal random basis vector ϕ_m. If the incoherence condition between ϕ and ψ are satisfied, then there is a high probability that y can be reconstructed perfectly if $M > K \log N$ measurements.

At Receiver side, for reconstruction of signal, convex optimization techniques are used [6].

Convex optimization then can be utilized as follows [6]:

$$\hat{s} = \arg\min \|s\|_p \text{ subject to } y = \psi \cdot \phi \cdot s \ \ \& \ \ \hat{x} = \psi \cdot \hat{s} \tag{3}$$

Where, $\|\bullet\|_1$ is the l_1-norm. The algorithm above is known as “Basis Pursuit” (BP) since a subset of the column vector of $\phi\psi$ is being determined.

Another efficient algorithm to solve CS is “orthogonal matching pursuit” (OMP) which can be formulated as follows [6]:

$$\hat{s} = \arg\min \|y - \phi \cdot \psi \cdot s\|_2 \text{ and } \|s\|_0 = K \tag{4}$$

Because of the time varying nature of speech signal, sensing and compressing are applied on a short duration of the signal. It is known that the perceptually significant features of spectral resonances and the harmonicity due to periodic excitation, are the most important and basic parameters in speech and audio [6]. Therefore, to explore sparsity of the speech signal, several alternative representation of a speech frame can be considered, such as

$$x = C^{-1}\theta_1 \tag{5}$$

$$x = F^{-1}\theta_2 \tag{6}$$

Eq. (5) gives representation of x in terms of DCT where C is the real valued transform matrix and θ_1 is the DCT coefficients [6]. Similarly, in Eq. (6), θ_2 corresponds to the DFT matrix F, which is complex valued. Hence, various transforms, such as Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT) with various wavelet bases, can be used to sparsify the speech signal. The best transform that provides higher sparsity index, *i.e.*, more Sparsity, can be selected by using Gini index as it provides the best measurement [10]. In this paper, DCT sparsity basis will be considered.

2.1 Sensing Matrix

2.1.1 Random Matrix

Independent and Identical Distributed (i.i.d.) Gaussian Matrix is generated by choosing all entries randomly and independently from normal distribution with mean zero and variance 1/n [10].

2.1.2 Bernoulli Matrix

By sampling I. I. D. entries from a symmetric Bernoulli distribution (P (Ai, j= ±1/√m) = 1/2) or other sub-Gaussian distribution [2].

2.1.3 Fourier Matrix

It has randomly selected M rows out of the N × N Fourier matrix. Computation complexity can be reduced by using the fast Fourier transform algorithms. It is used for time domain only [10].

2.1.4 Wavelet

They can be used as sparse basis and sensing matrices because of its inherent property of orthogonality among rows & columns. A combination of wavelets with DCT can give very good reconstruction results in less time [10].

2.1.5 Orthogonal Symmetric Toeplitz Matrices (OSTM)

They are easy to generate as only N numbers need to be stored. Both sampling and reconstruction are more efficient to implement. This matrix is generated using three sequences Binary, Ternary and PN Sequence and specially designed for speech signal.

2.1.5.1 Binary Sequence

A binary number is a number expressed in the binary numeral system, or base-2 numeral system, which represents numeric values using two digits: 0 and 1.

2.1.5.2 Ternary Sequence

Ternary (sometimes called trinary) is the base-3 numeral system that represents numeric values using the three digits: 0, 1 and 2.

2.1.5.3 PN Sequence

Pseudorandom binary sequences can be generated using "linear feedback shift registers".

Procedure for generating Orthogonal Symmetric Toeplitz Matrix using Binary, Ternary and PN Sequence:

An M × N sensing matrix based on OSTM can be constructed like this [14]:

1. Use a given sequence **s** of length N,
$$\sigma = [S_1, S_2 \ldots S_N] \quad (7)$$
2. And apply inverse FFT (IFFT) to the sequence to obtain g with length N.
$$g = \text{ifft}(\sigma) \quad (8)$$
3. Let the elements of g be the first row of OSTM, and follow the circulant property to construct the N × N matrix ϕ_N.
4. Choose M rows and normalize it by multiplying $\sqrt{N/M}$ to form the M × N sensing matrix ϕ_N. After the second step, it can be proved that the N × N matrix ϕ_N is orthogonal and Toeplitz.

Copyright ⓒ 2014 SERSC

By following this procedure Orthogonal Symmetric Matrix using given three sequences is generated.

3. Simulation

The experiment is conducted on a speech files taken from NOIZEUS database. Male file contains 22400 samples and female file has 20160 samples. The sampling rate is 8 KHz. This test is conducted on MATLAB with i3 Intel Core Processor Clock frequency at 2.53 GHz. The whole speech is divided into number of frames. Each frame contains 160 samples. Here, Orthogonal Gaussian matrix is taken as sensing matrix. Threshold value is found by following equation:

$$pos.threshold = mean(0 \le samples \le 0.05) \tag{9}$$

$$Neg.threshold = mean(0 < samples \ge -0.05) \tag{10}$$

For reconstruction of speech signal l_1-minimization and OMP optimization techniques are taken here. Here, Compression effect on speech by compressive sensing is tested by taking numbers of measurements 80, 100, and 120 for DCT, LPC and DFT are sparsity and for performance measurement above discussed parameters are taken.

3.1. Performance Matrix

Different Three performance metrics are used to quantify the compression techniques. Here, the comparison is done between original signal x[n] and reconstructed signal y[n] with different compression ratio (CR) .compression ratio is defined as ratio of M/N where M are the number of measurement taken for a frame and N are the number of samples present per frame. Following are the parameters based on that performance is evaluated:

3.1.1. Mean Square Error

For the original speech x[n] and the synthetic version y[n], with the range of the time index n covering the measurement interval, the MSE is defined by,

$$MSE = \frac{\sum_n (x[n] - y[n])^2}{n} \tag{11}$$

MSE shows the amount by which reconstructed speech differs from the original speech.

3.1.2. Signal to Noise Ratio (SNR)

Given the original speech x[n] and the synthetic version y[n], with the range of the time index n covering the measurement interval, the SNR is defined by,

$$SNR = 10\log_{10}\left(\frac{\sum_n x[n]^2}{\sum_n (x[n] - y[n])^2}\right) \tag{12}$$

3.1.3. Perceptual Evaluation Speech Quality (PESQ)

PESQ means perceptual evaluation of speech quality which is one of the most reliable methods to evaluate the performance of the Speech quality. It helps to find the degradation of the signal. It is calculated by using the subjective opinion score. The range of PESQ lies within 0.5 to 4.5, with the lower values interpreting as poor speech quality [11].

3.2 Results

Experiment is conducted on sp01.wav (Male File) and sp13.wav (Female file) with DCT basis and various Sensing Matrices.

Table 1. Comparison between Various Sensing Matrices (M=120)

DCT Basis M=120 (MALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	4.36E-05	14.32	3.30	1.83E-05	18.09	3.69
Bernoulli	4.31E-05	14.36	3.20	2.12E-05	17.45	3.67
Hadamard	4.36E-05	14.32	3.30	1.83E-05	18.09	3.69
Fourier	8.03E-05	11.66	3.09	2.33E-05	17.04	3.64
OSTM (Binary)	3.50E-05	15.26	3.28	1.73E-06	28.32	3.93
OSTM (Ternary)	3.09E-05	15.82	3.39	1.37E-06	29.35	3.94
OSTM (PN)	3.86E-05	14.84	3.50	2.45E-06	26.82	3.88

Table 2. Comparison between Various Sensing Matrices (M=100)

DCT Basis M=100 (MALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	1.02E-04	10.60	2.80	6.16E-05	12.81	3.21
Bernoulli	1.37E-04	9.35	2.85	6.51E-05	12.57	3.23
Hadamard	1.33E-04	9.46	2.64	1.24E-04	9.78	2.67
Fourier	1.83E-04	8.09	2.53	7.73E-05	11.82	3.18
OSTM (Binary)	4.89E-05	13.81	3.08	6.91E-06	22.31	3.79
OSTM (Ternary)	2.86E-05	16.15	3.01	4.70E-06	23.99	3.38
OSTM (PN)	3.45E-05	15.33	2.94	9.83E-06	20.78	3.33

Table 3. Comparison between Various Sensing Matrices (M=80)

DCT Basis M=80 MALE	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	2.63E-04	6.51	2.34	1.57E-04	8.74	2.65
Bernoulli	2.40E-04	6.90	2.45	1.48E-04	9.01	2.65
Hadamard	1.13E-03	0.19	1.74	5.91E-04	2.99	1.82
Fourier	4.43E-04	4.24	2.15	1.52E-04	8.89	2.70
OSTM (Binary)	5.16E-05	13.58	2.82	1.81E-05	18.14	3.27
OSTM (Ternary)	3.83E-05	14.87	2.69	1.22E-05	19.86	3.07
OSTM (PN)	3.76E-05	14.96	2.96	2.28E-05	17.13	3.13

 Copyright ⓒ 2014 SERSC

Table 4. Comparison between Various Sensing Matrices (M=120)

DCT Basis M=120 FEMALE	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	1.06E-04	13.37	2.94	7.12E-05	15.11	3.23
Bernoulli	1.06E-04	13.38	2.96	6.94E-05	15.22	3.30
Hadamard	7.72E-05	14.76	3.01	1.18E-04	12.91	2.96
Fourier	2.20E-04	10.21	2.66	7.30E-05	15.00	3.18
OSTM (Binary)	8.02E-05	14.59	3.12	1.39E-05	22.22	3.78
OSTM (Ternary)	5.19E-05	16.48	3.20	1.06E-05	23.38	3.79
OSTM (PN)	6.02E-05	15.84	3.04	1.39E-05	22.21	3.64

Table 5. Comparison between Various Sensing Matrices (M=100)

DCT Basis M=100 FEMALE	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	2.79E-04	9.17	2.55	1.85E-04	10.96	2.79
Bernoulli	2.63E-04	9.44	2.48	1.80E-04	11.09	2.79
Hadamard	2.16E-04	10.30	2.61	2.69E-04	9.34	2.66
Fourier	4.43E-04	7.18	2.32	1.69E-04	11.35	2.76
OSTM (Binary)	8.57E-05	14.31	2.92	4.19E-05	17.42	3.54
OSTM (Ternary)	5.90E-05	15.92	3.11	3.03E-05	18.82	3.59
OSTM (PN)	6.60E-05	15.44	3.10	4.23E-05	17.38	3.43

Table 6. Comparison between Various Sensing Matrices (M=80)

DCT Basis M=80 FEMALE	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
Random	7.10E-04	5.12	2.21	3.99E-04	7.62	2.42
Bernoulli	6.99E-04	5.19	2.17	3.56E-04	8.12	2.47
Hadamard	1.34E-03	2.36	1.79	1.17E-03	2.95	1.99
Fourier	9.52E-04	3.85	2.03	3.64E-04	8.03	2.47
OSTM (Binary)	9.42E-05	13.89	2.92	6.51E-05	15.50	3.30
OSTM (Ternary)	6.55E-05	15.47	3.10	4.58E-05	17.02	3.41
OSTM (PN)	8.35E-05	14.42	3.02	6.88E-05	15.26	3.18

Copyright ⓒ 2014 SERSC

From the tables, it is clear that OSTM (Ternary) matrix gives best results among all i.e. least MSE value and highest PESQ and SNR values among all sensing matrices. And all OSTM (Binary, Ternary and PN-Sequence) matrices give good quality results compared Random, Bernoulli, Hadamard and Fourier matrix. OSTM generated by PN Sequence give poor results compared to OSTM generated by Binary and Ternary Sequences. From all sensing matrices, Hadamard gives poor results compared to others. As numbers of measurements decreases, Mean square error increases and SNR and PESQ values decreases in all basis. L_1-minimization gives better result compared to OMP algorithm.

4. Conclusion

Compressive Sensing can be efficiently used in speech processing applications. Sampling, Compression and encryption is obtained using Sensing Matrix and Sparsity Domain Conversion. Simulation results show that Hadamard (Wavelet) gives poor results compared to all other sensing Matrix with DCT basis whereas; Orthogonal Symmetric Toeplitz Matrices (Binary, Ternary and PN-Sequence) give good results with DCT compared to others. From all Orthogonal symmetric Toeplitz Matrices, OSTM generated by PN Sequence give poor speech quality of reconstructed speech Signal compared to OSTM generated by Binary and Ternary Sequences for all measurements.

Acknowledgements

The authors are grateful to all their colleagues for their useful comments and cooperation on many topics related to this work. Also, we would like to thank our organization and staff members for the support and the facilities that have been provided to us.

References

[1] D. L. Donoho, "Compressive sensing. Department of statistics", Stanford University, **(2004).**
[2] E. J. Candes and M. B. Wakin, "An Introduction to Compressive sampling", IEEE signal Processing Magazine, vol. 25, Issue. 2, **(2008),** pp. 21-30.
[3] T. V. Sreenivas and W. B. klejin, "Compressive sensing for sparsely excited speech signals", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, **(2009),** pp. 4125-4128.
[4] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jenson and M. Moonen, "Retrieving Sparse patterns using a compressed sensing framework: Applications to speech coding Based on sparse linear prediction", IEEE Signal processing letters, vol. 17, Issue. 1, **(2010),** pp. 103-106.
[5] H.-N. Lee, "Introduction to compressive sensing. Lecture notes, **(2011),** pp. 26-29.
[6] T. S. Gunawan, O. O. Khalifa, A. A. Shafie and E. Ambikairajah, "Speech compression using compressive sensing on multicore system", 4th International Conference on Mechatronics (ICOM), Kuala Lumpur, **(2011),** pp. 1-4.
[7] Y. Wang, Z. Xu, G. Li, L. Chang and C. Hong, "Compressive Sensing Framework for Speech Signal Synthesis Using a Hybrid Dictionary", 4th International Congress on Image and Signal Processing (CISP), Shanghai, vol. 5, **(2011),** pp. 2400-2403.
[8] A. Sabir, "Compressive sensing for speech signals in mobile system", M.S. Thesis. The University of Texas, **(2011).**
[9] L. A. kassim and T. S. Gunawan, "Evaluation of sparsifying algorithm for speech signals", International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, (**2012)**, pp. 308-313.
[10] V. Abrol, P. Sharma and S. Budhiraja, "Evaluating performance of compressed sensing for speech signal", IEEE 3rd International advance computing conference (IACC), Ghaziabad, (**2013)**, pp. 1159-1164.
[11] "The PESQ Algorithm as the Solution for Speech Quality Evaluation on 2.5G and 3G Network Technical Paper", URL: http://www.cn.ascom.com/cn/pesq-3g.pdf.
[12] "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms", URL: http://ecs.utdallas.edu/loizou/speech/noizeus.

 Copyright ⓒ 2014 SERSC

[13] S. Desai and N. Nakrani, "Compressive Sensing in speech processing: A survey based on sparsity and sensing matrix, IJETAE, vol. 3, issue 12, (**2013)**.
[14] K. Li, C. Ling and L. Gan, "Deterministic compressed sensing matrices: Where Toeplitz meets Golay", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **(2011)**, pp. 3748-3751.

Authors

Miss. Siddhi Desai, Author has completed B.E.in Electronics and communication in 2012 from Gujarat Technological University (GTU) and currently doing M.Tech in Electronics and Communication from Chhotubhai Gopalbhai Patel Institute of Technology (CGPIT), Bardoli from Uka Tarsadia University, INDIA.

Mr. Naitik Nakrani, Author has completed B.E.in Electronics and communication from DDIT University, Gujarat and M.E in Electronics and Communication from SCET, Surat from Gujarat Technological University, INDIA. Currently Author is working as Assistant professor in Chhotubhai Gopalbhai Patel Institute of Technology (CGPIT), Bardoli from Uka Tarsadia University, INDIA.

Copyright ⓒ 2014 SERSC

 Copyright ⓒ 2014 SERSC