ISSN:0975-9646

Evaluating Performance of Compressive sensing for speech signal with Combined Basis

Siddhi Desai

[1]*PG Student, Electronics & Communication Department, CGPIT, Bardoli, UTU, INDIA.*

***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 approach called combined basis that is made by separating voiced and unvoiced parts and applying different basis for both parts from given speech and shows detailed comparison of them with LPC basis and orthogonal gaussian matrix applied on 8 KHz sampled speech signal. Also it shows improved results of Combined DCT and LPC basis compared to LPC and Combined DFT and LPC basis. Performance of these basis has been compared with Mean square error, Signal to noise ratio and Perceptual Evaluation of Speech Quality (PESQ) parameters.**

***Keywords*— Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Linear Prediction Coding (LPC), Orthogonal Gaussian Matrix**

I. 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 compression with some rigorous evaluation. Therefore, the objective of this paper is to explore a new idea on speech compression based on compressive sensing by considering new approach of taking combined basis and its performance is compared with LPC basis by quality assessment parameters like Mean Square Error (MSE), Signal to Noise Ratio (SNR) and Perceptual Evaluation of Speech Quality (PESQ).

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. In section III analysis of Compressive sensing for speech compression application is done with different sparsity basis. The conclusion is given in section VI.

II. COMPRESSIVE SENSING BASICS

The basic principle of Compressive Sensing is shown in Fig.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].

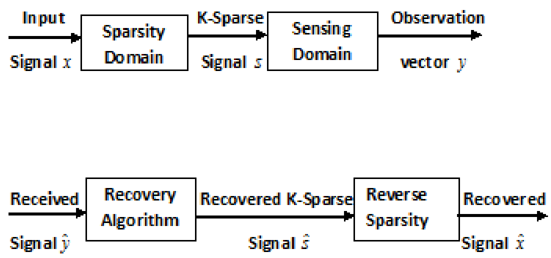

Fig. 1: Block Diagram

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

Let $x \in R^N$ be the speech signal and let $\psi = [\psi_1, \psi_2 . . \psi_N]$ be the basis vectors spanning R^N.

The speech signal is said to be sparse if,

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

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:

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

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]:

$$\hat{s} = \operatorname{argmin} \|s\|_p \text{ subject to } y = \psi \cdot \phi \cdot s \;\&\; \hat{x} = \psi \cdot \hat{s} \qquad (3)$$

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]:

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

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

$$x = C^{-1}\theta_1 \qquad (5)$$

$$x = F^{-1}\theta_2 \qquad (6)$$

$$x = A^{-1} \cdot r = H \cdot r \qquad (7)$$

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. Eq. (7) gives representation of x in terms of LP residual vector. Where, r= residue vector and. H is the inverse matrix of A (Matrix that performs the whitening of the signal, constructed from the coefficients of the predictor a of order P) and it is commonly referred to as the synthesis matrix that maps the residual representation to the original speech domain and $\psi = H$ [4]. Hence, various transforms, such as Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and Linear Prediction coding can be used to sparsify the speech signal. In this paper, we try to apply one new approach. In this approach, First voiced and unvoiced part is separated using short time energy calculation of each frame. For both voiced and unvoiced part, different basis are chosen. For example, DFT for voiced part and LPC for unvoiced part.

III. 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. By Short time energy calculation voiced and unvoiced part is separated. For separating voiced and unvoiced part threshold value 1 is taken. Here, Orthogonal Gaussian matrix is taken as sensing matrix. Four combined basis are taken here for analysis, those are listed below:

1. DCT (Voiced) and LPC (Unvoiced)
2. DCT (Unvoiced) and LPC (Voiced)
3. DFT (Voiced) and LPC (Unvoiced)
4. DFT (Unvoiced) and LPC (Voiced)

Threshold value is found by following equation:

$$pos.threshold = mean(0 \le samples \le 0.05) \qquad (8)$$

$$Neg.threshold = mean(0 < samples \ge -0.05) \qquad (9)$$

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 LPC and four above listed basis. And for performance measurement below discussed parameters are taken.

A. *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:

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,

$$MSE = \frac{\sum_n (x[n] - y[n])^2}{n} \qquad (10)$$

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

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,

$$SNR = 10\log_{10}\left(\frac{\sum_n x[n]^2}{\sum_n (x[n] - y[n])^2}\right) \qquad (11)$$

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].

B. *Results*

Experiment is conducted on sp01.wav (Male File) and sp13.wav (Female file) with LPC and four combined basis listed above. Following tables show obtained results:

Table 1: Comparison between LPC and Combined Basis (M=120)

Sparsity Basis M=120 (MALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	3.28E-04	5.55	2.61	1.64E-04	8.56	2.93
DFT(V)+LPC(U)	3.76E-04	4.96	2.59	2.54E-04	6.66	2.86
DCT(V)+LPC(U)	2.44E-04	6.84	2.69	1.15E-04	10.08	3.02
DFT(U)+LPC(V)	4.46E-04	4.21	2.14	4.04E-04	4.65	2.17
DCT(U)+LPC(V)	1.21E-04	9.86	3.27	6.09E-05	12.86	3.55

Table 2: Comparison between LPC and Combined Basis (M=100)

Sparsity Basis M=100 (MALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	6.86E-04	2.35	2.29	3.34E-04	5.47	2.59
DFT(V)+LPC(U)	6.30E-04	2.72	2.30	3.84E-04	4.87	2.53
DCT(V)+LPC(U)	4.64E-04	4.05	2.40	2.64E-04	6.50	2.58
DFT(U)+LPC(V)	5.75E-04	3.11	2.00	4.54E-04	4.14	2.03
DCT(U)+LPC(V)	2.81E-04	6.23	2.76	1.33E-04	9.48	3.13

Table 3: Comparison between LPC and Combined Basis (M=80)

Sparsity Basis M=80 (MALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	1.04E-03	0.54	2.15	4.67E-04	4.02	2.46
DFT(V)+LPC(U)	8.31E-04	1.51	2.14	4.78E-04	3.92	2.41
DCT(V)+LPC(U)	7.83E-04	1.77	2.11	3.91E-04	4.79	2.39
DFT(U)+LPC(V)	7.14E-04	2.17	1.84	5.25E-04	3.51	1.82
DCT(U)+LPC(V)	4.52E-04	4.16	2.47	2.31E-04	7.07	2.68

Table 4: Comparison between LPC and Combined Basis (M=120)

Sparsity Basis M=120 (FEMALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	8.69E-04	4.24	2.01	4.27E-04	7.34	2.36
DFT(V)+LPC(U)	1.15E-03	3.03	2.08	1.01E-03	3.60	2.27
DCT(V)+LPC(U)	3.18E-04	8.61	2.48	1.54E-04	11.76	2.78
DFT(U)+LPC(V)	8.87E-04	4.15	2.03	5.71E-04	6.07	2.22
DCT(U)+LPC(V)	6.68E-04	5.39	2.25	3.38E-04	8.35	2.56

Table 5: Comparison between LPC and Combined Basis (M=100)

Sparsity Basis M=100 (FEMALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	1.38E-03	2.23	1.86	7.32E-04	4.99	2.13
DFT(V)+LPC(U)	1.30E-03	2.48	1.99	1.10E-03	3.22	2.13
DCT(V)+LPC(U)	5.43E-04	6.29	2.22	2.95E-04	8.94	2.42
DFT(U)+LPC(V)	1.29E-03	2.53	1.88	8.28E-04	4.46	2.09
DCT(U)+LPC(V)	1.12E-03	3.16	2.04	6.16E-04	5.74	2.37

Table 6: Comparison between LPC and Combined Basis (M=80)

Sparsity Basis M=80 (FEMALE)	Reconstruction Algorithm					
	OMP			L_1-Minimization		
	MSE	SNR (db)	PESQ	MSE	SNR (db)	PESQ
LPC	2.09E-03	0.43	1.67	1.08E-03	3.30	1.93
DFT(V)+LPC(U)	1.59E-03	1.63	1.88	1.26E-03	2.63	2.00
DCT(V)+LPC(U)	9.57E-04	3.83	2.06	5.37E-04	6.33	2.24
DFT(U)+LPC(V)	1.88E-03	0.90	1.68	1.10E-03	3.21	1.89
DCT(U)+LPC(V)	1.82E-03	1.04	1.75	9.35E-04	3.93	2.06

From the above tables it is observed that, Combined DFT+LPC give poor results compared to LPC and Combined DCT+LPC for both reconstructions Algorithm. Instead of LPC, Combined LPC+DCT can prove better choice as basis for speech. For all basis, as numbers of measurements decreases, Mean square error increases and SNR and PESQ values decreases. L_1-Minimization gives better reconstruction compared to OMP for all choices of measurements.

IV. CONCLUSIONS

Compressive sensing theory can be efficiently used in speech processing applications. Due to the Sensing matrix and Sparsity domain conversion of signal in compressive sensing, sampling, compression and encryption is obtained. Compressive sensing theory can be efficiently used for both Male and Female voice if sparsity basis are properly chosen. Combined DFT+LPC give poor results compared to LPC and Combined DCT+LPC for both reconstructions Algorithm. Instead of LPC, Combined LPC+DCT can prove better choice as basis for speech. For all basis, as numbers of measurements decreases, Mean square error increases and SNR and PESQ values decreases. L_1-Minimization gives better reconstruction compared to OMP for all choices of measurements.

ACKNOWLEDGMENT

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] David L. Donoho. 2004. Compressive sensing. Department of statistics, Stanford University.

[2] Emmanuel J. Candes and Michael B. Wakin, "An Introduction to Compressive sampling." IEEE signal Processing Magazine, Vol.25, Issue.2, 2008. 21-30.

[3] T. V. Sreenivas and W. Bastiaan klejin, "Compressive sensing for sparsely excited speech signals." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, 2009. 4125 – 4128.

[4] Daniele Giacobello, Mads Graesboll Christensen, Manohar N. Murthi, Soren Holdt Jenson, Marc 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, 2010.103-106.

[5] Heung-No lee. 2011. Introduction to compressive sensing. Lecture notes. 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, 2011.1-4.

[7] Yue Wang, Zhixing Xu, Gang Li, Liping Chang and Chuanrong Hong. "Compressive Sensing Framework for Speech Signal Synthesis Using a Hybrid Dictionary." 4th International Congress on Image and Signal Processing (CISP), Vol.5, Shanghai, 2011. 2400 – 2403.

[8] Ahmed Sabir. 2011. Compressive sensing for speech signals in mobile system. M.S. Thesis. The University of Texas.

[9] Liban A. kassim and T. S. Gunawan, "Evaluation of sparsifying algorithm for speech signals." International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, 2012.308 – 313.

[10] Vinayak Abrol, Pulkit Sharma and Sumit Budhiraja,"Evaluating performance of compressed sensing for speech signal." IEEE 3rd International advance computing conference (IACC), Ghaziabad, 2013.1159-1164.

[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

[13] Compressive Sensing in speech processing: A survey based on sparsity and sensing matrix". IJETAE vol. 3 issue 12. December-2013.