INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

COMPRESSIVE SENSING IN SPEECH FROM LPC USING GRADIENT PROJECTION FOR SPARSE RECONSTRUCTION

Bhogayta Chetankumar Dilipbhai[1]*, Suneet Betrabet[2]
[1*2]School of Electronics Engineering (SENSE), Vellore Institute of Technology University, Chennai-600127, Tamilnadu, India
Correspondence Author: bhogayta.chetankumar2013@vit.ac.in

Keywords: Compressive Sensing, GPSR, IDCT, LPC.

Abstract
This paper presents compressive sensing technique used for speech reconstruction using Linear predictive coding because the speech is more sparse in LPC. DCT of a speech is taken and the DCT points of sparse speech are thrown away arbitrarily. This is achieved by making some point in DCT domain to be zero by multiplying with mask functions. From the incomplete points in DCT domain, the original speech is reconstructed using compressive sensing and the tool used is Gradient Projection for Sparse Reconstruction. The performance of the result is compared with direct IDCT subjectively. The experiment is done and it is observed that the performance is better for compressive sensing than the DCT.

Introduction
Compressive Sensing is a signal processing technique for reconstructing a signal,[1] by finding solution to undetermined linear equations. Compressive sensing is a new type of sampling theory, which predicts that sparse signals and images can be reconstructed from what was previously believed to be incomplete information**.**

As our modern world technology-driven civilization acquires and exploits ever-increasing amounts of data, 'everyone' now knows that most of the data we acquire 'can be thrown away' with almost no perceptual loss. Instead of sensing all information and throwing away later. Compressive sensing is a method to sense only the required information. Compressive Sensing sample the signal with lower rate than sampling frequency fs. This Compressive measurement can be small and still contain all the useful information. Compressive sensing is used in mobile phone camera sensor, MRI scanning sessions, holography, astronomy etc.

The fundamental mathematical formulation of CS [1-2] is discussed here to build a CS framework for our requirement. Let us assume that a 1-D signal x having N samples is measured in transform domain given by a linear mapping

$$y = Ax \qquad (1)$$

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

$$x = A^{-1}y \qquad (2)$$

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 [4], if x is S sparse and if the following equation (3) is also obeyed.

$$K \geq C.S.logN \qquad (3)$$

Where, C is constant

From a general viewpoint, sparsity and, more generally, compressibility has played a fundamental role in many fields of science. Sparsity leads to efficient estimations; for example, the quality of estimation by thresholding algorithms depends on the sparsity of the signal we wish to estimate. Sparsity leads to efficient compression; for example, the precision of a transform coder depends on the sparsity of the signal

Application of CS to speech and audio is not straight, since the signals constitute a very large class of production mechanisms, emphasizing different characteristics of the signal at different times. The domain in which their sparsity can be exploited is also not clear and their degree of sparsity. The perceptual properties of the reconstructed signal and the computational constraints also become important for a practical application, since the basic CS formulation is very computation intensive. In this paper, we show that recovery is possible from sub-Nyquist rate CS of speech the linear system, using Linear predictive Coding (LPC).

There are several algorithms implemented using compressive sensing [2].In our present work we had taken LPC as transformation [3] and compared the performance for various speech applied in LPC domain. The performance is also compared with direct Inverse DCT.

Methodology
GPSR (Gradient projection of sparse reconstruction)
Many problems in signal processing and statistical inference involve finding sparse solution to under-determined or ill-conditioned linear system of equations. GPSR algorithm is one of the best techniques to reconstruct the sparse signal. A standard

INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

approach consists in minimizing an objective function which includes a quadratic (squared) error term combined with a sparseness-inducing regularization term. There has been considerable interest in solving the convex unconstrained optimization problem

$$\underbrace{\min}_{x} \frac{1}{2}\|y - Ax\|_2^2 + \tau\|x\|_1 \qquad (4)$$

Where $x \in \mathbf{R}^n$, $y \in \mathbf{R}^k$, A is an $k \times n$ matrix, τ is a nonnegative parameter. [3]GPSR 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

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

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.

In this section an alternative interpretation of above transformation matrix A and B are considered. These matrices constitute the kernel functions (e.g twiddle factor in DFT) of the transform selected and for our convenience of implementation let us assume them as an operator A and B that process on x to produce the transformed output. Similary the inverse operators A^{-1} and B^{-1} are also available. Large scale implementation of CS algorithm requires implementation of these operators which are used to iteratively solve the optimization problem in (5). This helps the implementation to avoid representation of complex and large A and B matrices and also helps to use the existing fast blocks to determine the transforms.

Linear predictive coding (LPC)

LPC methods are the most widely used in speech coding, speech synthesis, speech recognition, speaker recognition and verification and for speech storage LPC methods provide extremely accurate estimates of speech parameters, and does it extremely. Efficiently basic idea of Linear Prediction: current speech sample can be closely approximated as a linear combination of past samples. Speech can be modeled as the output of a linear, time-varying system, excited by either quasi-periodic pulses or noise; assume that the model parameters remain constant over speech analysis interval[6]. LPC provides a robust, reliable and accurate method for estimating the parameters of the linear system.

For efficient coding or storage of speech, speech signals are often modeled using parameters of the presumed vocal tract shape generating them. In such modeling, it is important to analyze signals accurately and quickly.[6] One of the best techniques for analysis of certain physical signals is linear predictive coding (LPC). Signals produced by a relatively slowly-varying linear filtering process are most suitable for LPC, especially, if the filter is excited by infrequent, brief pulses. In particular, a signal's spectral magnitude, as opposed to phase, is widely used. The future values of such signals are well estimated by a linear predictor, which outputs values based on a linear combination of previous signal values.

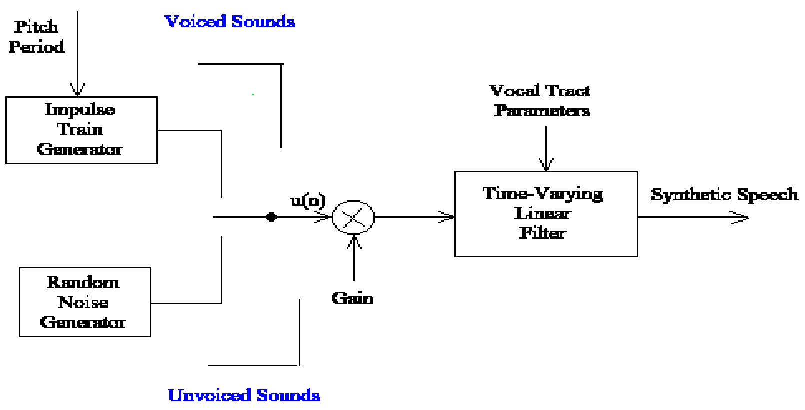


Fig. 1 Speech Synthesis model based on LPC

INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

Now, the selection of sparsifying operator B is the question in the case of compressive sensing of speech and the previous works are to define this B and an approach to solve eq (5). It is shown [6] 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, MFCC analysis that are used in the well known speech compression methods. This statement stands good for most of voiced segment of speech. The speech production model assume the vocal tract to be linear time invariant system momentarily for the production of every phonetics and the speech output is given by

$$x = e * h \qquad (6)$$

Where, as $*$ stands for convolution, h is the impulse response of the vocal tract which is a IIR filter and e is the input excitement signal assumed to be very sparse in nature for voiced phonetics and as noise for unvoiced phonetics. The popular Code Excited Linear Prediction (CELP)[6] compression technique use this h represented as LPC coefficients a which constitute the numerator of an all pole IIR system .A dictionary of codebook that comprise of representative vectors of various a is trained from a very long speech unit. As discussed in previous section let us represent (6) as

$$x = H.e \qquad (7)$$

Where, H is an operator that performs like an IIR filter. As shown in [3] we can observe that there are L possible operators in the place of H assuming the code book dictionary consists of L vectors (a typical value of 128) and (7) can be rewritten as

$$x = H_i.e \ \ for\ 1 < i < L \qquad (8)$$

This leads to L fold increase in optimization steps as now the objective function is given by

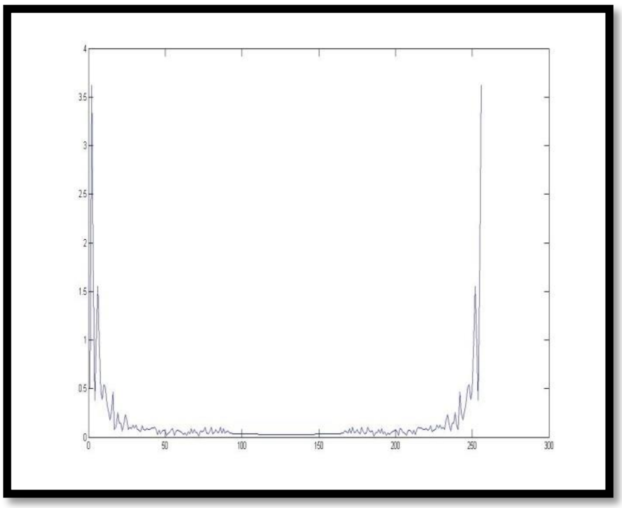
$$\underbrace{min}_{x,i} \frac{1}{2}\|y - AH_i e\|_2^2 + \tau\|e\|_1 \qquad (9)$$

Discrete cosine ransform (DCT)

One of the appliance bases for compressive sensing theory is signal sparse. Classic sparse transformations discrete cosine transform (DCT), discrete Fourier transformation (DFT) and discrete wavelet transformation (DWT). Really DCT is most near the K-L transformation of best transformation performance. Thus it's usually utilized in speech and image process. A technique of partial two-dimensional DCT combined CS utilized in the speech coding & decoding systems. DCT has the property of unifying energy to low frequency. The original speech is rotten into high waveband and low waveband once DCT transform, high waveband is thought-about to be sparse as, and however the low waveband coefficients don't seem to be sparse.

Numerical simulation

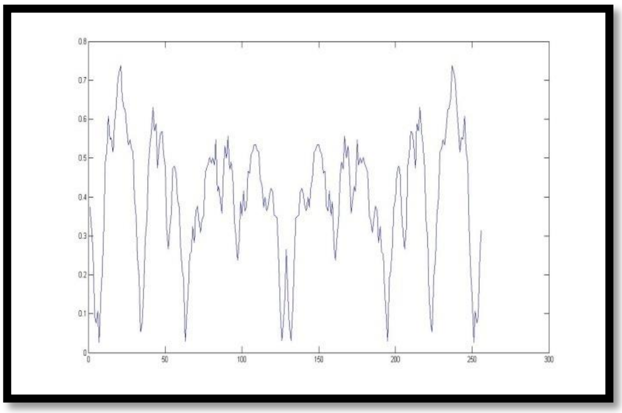
1. The speech which contains some part of total speech is shown in fig(2).That denoted by I.



Fig(2): Original speech signal(I)

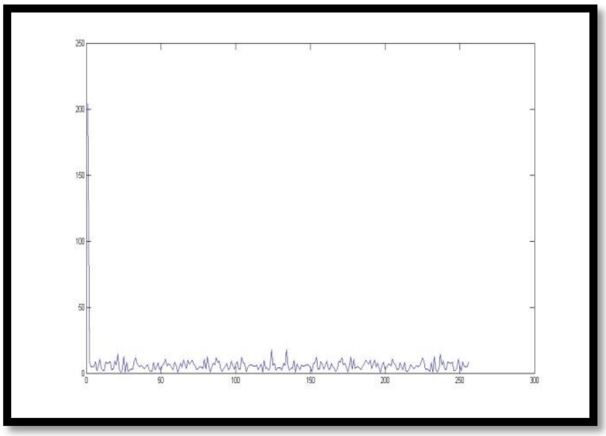
INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

2. Than DCT apply to I and we get R. It is shown in fig(3).



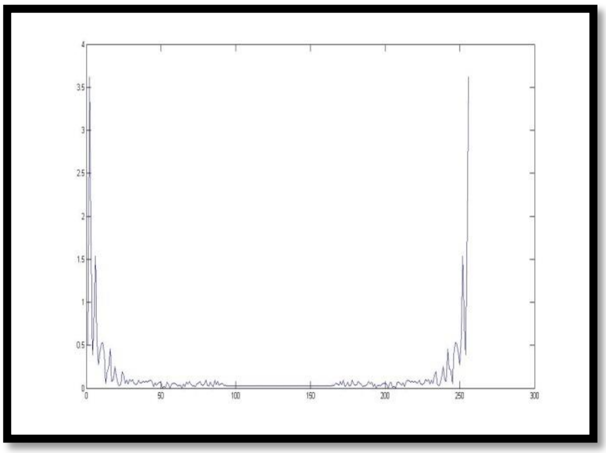
Fig(3): DCT speech signal(R)

3. Some random mask function is taken, which is shown in fig(3) and is multiplied with R and we get G.



Fig(4):Random mask function(G)

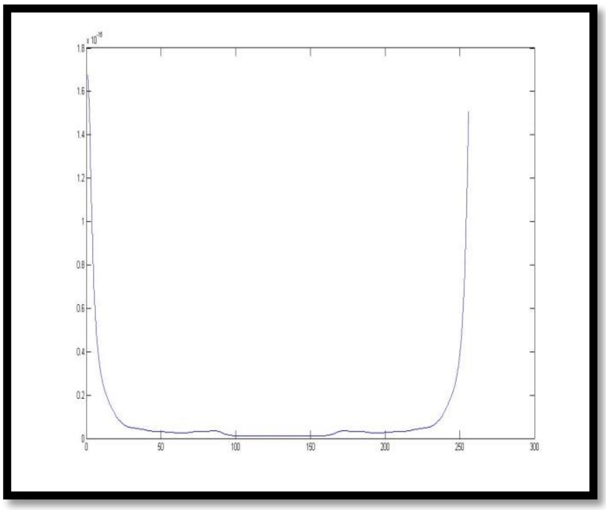
4. For the incomplete DCT information , compressive sensing is applied using GPSR to get reconstructed speech signal. It can be seen in fig(5).



Fig(5):Reconstructed using compressive sensing

INTERNATIONAL JOURNAL OF RESEARCH SCIENCE & MANAGEMENT

5. The reconstructed speech by direct IDCT is shown in fig(6). If we make comparison between fig(5) and fig(6) it can be seen that the result shown by the fig(5) is better than the result shown in fig(6).



Fig(6):Reconstructed using IDCT

Result analysis

By observing the figures we can make following interferences. For mask in which random points are thrown the reconstructing using compressive sensing method is subjectively better than direct IDCT method.

Conclusion

An analytical study, adopting compressive sensing technique and direct Inverse discrete cosine transform methods to reconstruct a sparse image in case of a defined mask function, results identical. However, compressive sensing technique adopted for sparse image reconstruction proves to be of higher prudence to random mask functions on comparison to direct Inverse discrete cosine transform method.

Acknowledgement

The project is result of 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 our school faculties and facilities provided by VIT management.

References

1. T.V. Sreenivas1 and W. Bastiaan Kleijn, “Compressive Sensin For Sparsely Excited Speech Signals,” IEEE Transl, 2009.
2. Candès.E.J., & Wakin.M.B,"An introduction to compressive sampling."IEEE Signal Proces Mag, 2009.
3. M´ario A. T. Figueiredo, Robert D. Nowak, Stephen J. Wright, “Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems”,IEEE Communication Magazine 2007.
4. Christenseny.M.G, Østergaardz.J & Jensenz.S.H. "On Compressed Sensing And It’s Application To Speech And Audio Signals," IEEE Conference Publications in Statistical Signal Processing, 2009.
5. Prabhakar.Telagarapu, V.Jagan Naveen, A.Lakshmi..Prasanthi, G.Vijaya Santhi.' Image Compression Using DCT and Wavelet Transformations 'International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 3, September, 2011.
6. G Sawant, A Singh, K Kadam, R Mazarello, P Dumane," Speech Synthesis using LPC", International Conference and Workshop on Emerging Trends in Technology (ICWET 2010)