Higher Order Adaptive Compressed Sensing of Speech Signal

Priyanka. M *P. Venkateswarlu* *G. V. S. Manoj kumar*

Abstract

Compressed sensing (CS) is an emerging signal acquisition theory that directly collects signals in a compressed form if they are sparse on some certain basics. This paper focuses on the realization of CS on speech signals. Observing that different kind speech frames have different intra-frame correlations, we propose a frame-based adaptive compressed sensing frame work. For speech signals, which applies adaptive projection matrix? Experimental results show significant improvement of speech reconstruction quality by using such adaptive approach against using such adaptive approach against using traditional non-adaptive projection matrix.

Key words-higher order, compressed sensing, speech signal, orthogonal matching pursuit

I. INTRODUCTION

In any wireless communication system, one of the most desirable features user want is high speed data transfer. However, most of the current wireless systems are limited mainly by their respective bandwidths. The project is low cost. It is the research premise is that signals have sparse representations where most of the coefficients are zeros or very small on some certain basis. CS is a new sampling and recovery model has emerged in recent years with much excitement and many papers deals it impact on different areas of research such as natural-image acquatization, remote sensing, cognitive radio and medical imaging etc. CS combines both V signals acquisition and dimensionality reduction in to a single step potentially re reducing computational and hardware burdens on sensing devices. The promise of CS to lower sampling cost and signal recovery. This makes CS quite attractive in many questions.

Conventional speech signal acquisition approaches follow Shannon's celebrated sampling theory: for accurate representation of a signal by its time samples, the sampling rate must be at least twice the maximum frequency presented by its time samples, the sampling rate must be at least twice the maximum frequency presented in the signal (the so called Nyquist rate). The dominant approach currently used for speech according to Nyquist rate and then to eliminate redundancy using various compression schemes such as LPC. The speech signal's high-frequency component which is presented just instantaneously will lead to unnecessary large amount of data, which is inconvenient for signal collection storage, processing and transmission. According to the characteristics of human auditory system and the correlation between speech time samples, it is recognized that speech signals are highly compressible, which provides the basis for CS algorithms. For example, the speech signals are considered sparse in the discrete cosine transform (DCT) domain. The speech signal acquisition approach based on Cs theory must be a revolutionary change in the speech signal processing filed.

Most work in CS research focus on random projection matrix which is constructed by considering only the signals scarcity rather than other properties. In other word, the construction of projection matrix is non-adaptive. Observing that different kind speech frames have different intra-frame correlations, this paper proposes a frame-based adaptive compressed sensing frame work for speech signals, which applies adaptive projection matrix. To do so, we compare neighboring frames to estimate their intra-frame correlation, classify every frame into different categories, and adjust the number of projections accordingly. The experimental results show that the adaptive projection matrix can significantly improve the speech reconstruction quality.

The rest of this paper organized as following: in section II, basic principles of CS are described. Section III describes our proposed adaptive frame based frame work for compressive speech sampling. Experimental results are shown in section IV, while conclusion is made in section V.

II COMPRESSED SENSING THEORY BACK GROUND

A.Basic Principle of compressed Sensing

Cs is a novel sensing paradigm which aims at reconstructing a sparse or compressible signal from its compressed measurements [1-2]

Let $x \epsilon R^N$ be a real valued signal of length N. we say that x is K-sparse on orthogonal sparse basis $\Psi \epsilon R^{NXN}$, if x can be represented by only K(K<<N) nonzero projections on Ψ . Thus the signal x can be written as

$$x = \Psi \qquad (1)$$

Where $\alpha = \Psi^T x$ is a length N coefficient vector with K nonzero elements.

The sampling scheme of Cs projects the signal on the projection matrix $\Phi = \epsilon R^{MXN}$ (M<N) where Φ and Ψ are incoherent. The CS sampling process can be expresses as:

$$y = \Phi \mathrm{x} = \Phi\Psi = \Xi\alpha \qquad (2)$$

Where $y \epsilon R^M$ is called the measurement vector of the original signal. As the number of measurement M is sometimes much smaller than the length of the signal N, the goal of signal compression can be achieved.

At CS decoder, the reconstruct algorithm tries to find the sparsest solutionα.

$$Min\ ||\alpha||_0 \qquad s.t\ \ \Xi\alpha=y \qquad (3).$$

Where $||\ ||_0$ is the l^0 norm, counting the nonzero entries of a vector. Then recovers the signal from the optional solutionα^* :

$$x^* = \Psi\alpha^* \qquad (4).$$

B.Some Popular Reconstruction Algorithms

Problem (3) is considered NP-hard so various sub-optional solutions have been proposed. Currently, two most popular approaches are basis pursuit (BP)[3] and orthogonal matching pursuit (OMP) {4}. BP seeks representations that minimize the l^1 -norm of the coefficients, which reduces signal representations to linear programming:

$$Min\ ||\alpha||_1 \qquad s.t\ \ \Xi\alpha=y \qquad (5).$$

OMP is a fast greedy algorithm that iteratively builds up a signal representation by selecting the atom that exactly maximally improves the representation at each iteration.

Because OMP is easily implemented and converges quickly, in this paper, we use OMP to test the reconstruction quality of speech signals.

III. ADAPTIVE SPEECH COMPRESSED SENSING

We explore the intra-frame correlation of speech signals to achieve efficient sampling. Note that different kind speech signals may have different intra-frame correlations, we propose a frame-based adaptive CS framework that uses different sampling strategies in different kind speech frames.

A. The Frame-based Adaptive CS framework

Each speech sequence is divided into non-overlapping frames of size1× n , and all frames in a speech sequence are processed independently. In our work, we initialize the projection matrix by Gaussian random matrix Φ , which has been proven to be incoherent with most sparse basises at high probability [1].

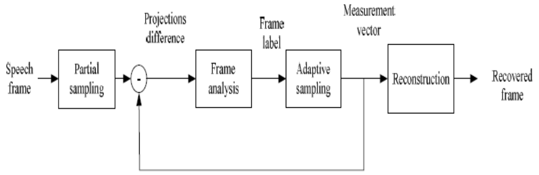

Figure 1. The frame-based adaptive CS framework for speech

As shown in Figure 1, for each frame in a speech sequence, we first collect a small number of projections, and compare it with the projections collected for the previous frame. Based on the comparison results, we estimate the correlation between these two frames, and classify the correlation into different categories. We then adjust the sampling strategy according to the correlation type and collect different number of samples for the current frame. The next sections *discuss details of each step in the above framework.*

B.Partial Sampling and Frame Analysis

For the current t th frame original speech signal, we represent it as x_t. Its previous frame t −1 is represented using x_{t-1} . The difference between x_t and x_{t-1} reflects the correlation between the two neighboring frames, and can be used to classify the correlation. Since $x_t - x_{t-1}$ is not available at the sampling stage. We use the collected measurements to estimate the correlation instead. The same projection matrix Φ is applied to all frames in the partial sampling stage, and we have where y_t and y_{t-1} are the projection vectors of x_t and x_{t-1} respectively. As each sample in$y_t - y_{t-1}$ is a linear combination of $x_t - x_{t-1}$, the difference between the two projection vectors also reflects the Intensity changes in the two frame. Therefore, we can estimate the amount of intensity changes in the two frames using only a small number of projections [6].

Let Φ_{M0} be a matrix containing the first M_0 rows of the Gaussian random matrix Φ . For the current frame t , we first use Φ_{M0} to collect M_0 measurements $y_t^{M0} = \Phi_{M0}\, x_t$ in the partial sampling stage. Then, we compare it with the first M_0 measurements in y_{t-1} and calculate the difference $y_t^d = y_t^{M_0} - y_{t-1}^{M_0}$. In the frame analysis module, given y_t^d, we calculate its l^2 norm normalized by M_0 and compare with two thresholds T_1 and T_2 ($T_1 < T_2$). If $||y_t^d||/ M_0 \leq T_1$, the current frame is almost the same as its previous frame. We consider the two neighboring frames may be both surd and label the intra-frame correlation as surd vs. surd. If 1 $T_1 \leq ||y_t^d||/ M_0 \leq T_2$, it indicates that these two neighboring frames undergo small changes. In this situation, the two neighboring frames may be both sonant at high probability and the intra-correlation is labeled as sonant vs. Sonant. If $||y_t^d||/ M_0 > T_2$, the two frames are significantly different from each other, which is most likely due to the change of the frame type, and we label the correlation as surd vs. sonant.

C. Adaptive Sampling

Depending on their classified intra-frame correlation types, different number of projections is used for the speech frames. We consider the frame as surd frame if its intra-frame correlation type is surd vs. surd. A surd frame contains the least new information in the speech. Thus, the M_0 measurements $y_t^{M_0}$ collected in the partial sampling stage are sufficient and we do not need additional sampling. That is $y^t = y_t^{M_0}$. When its intra-frame correlation is sonant vs. sonant, the frame is considered as sonant and contains some new information, which requires more measurements to be collected. For such frames, we collect M_1 ($M_1 > M_2$) measurements. We use the (M_0 +1)th to the M_1 th rows of the Gaussian random matrix Φ and combine with $y_t^{M_0}$ to generate the final projection vector y_t . The frames that experience large changes must contain the most new information. Therefore, we collect a total of M_2 ($M_1 > M_2$) measurements (including the initial M_0 measurements) during the sampling process. The total projection matrix is the first M_2 rows of the Gaussian random matrix Φ .

D.Reconstruction

After the sampling process, the entire speech sequence can be reconstructed frame by frame. Speech frames have adaptive number of measurements according to their intra-frame correlations and during the reconstruction, the CS projection matrixes is of adaptive size. In this paper, we use OMP as the Reconstruction algorithm.

IV. EXPERIMENTAL RESULT

To compare the performance of our proposed adaptive CS frame and the convential non-adaptive CS, we conduct some experiments. In our experiment, the testing speeches are chosen from CASIA Chinese speech library which is built by the China analysis institute of automation. The testing speech corpus consists of 200 utterances of Mandarin Chinese speech spoken by four speakers (two men and two women) and is 16kHz sampled and 16 bits quantized for each sample. Adaptive CS and CS sampling and reconstruction are performed frame by frame, with a frame length of N=320 samples. We use T_1 = 0.08 and T_2 = 0.4 , which is tested through a great number of experiments. Both the average-frame signal-to-noise ratio(AFSNR) and Mean Opinion Score(MOS) are carried to evaluate the performance of the frame-based adaptive CS frame with the non-adaptive CS. AFSNR is used to evaluate the reconstruction quality of speech signal:

$$AFSNR = \frac{1}{k}\sum_{k=1}^{k} 10\log(\frac{\|x_k\|_2^2}{\|x_k - \widehat{x_k}\|_2^2}) \quad ..(6)$$

Where K is the total frame number of a speech sequence; x_k and $\hat{x}_k$ represent the k th frame speech and the k^{th} h frame reconstructed speech. Under different compressed ratio, which is defined as r = M / N , the following test results are obtained based on the proposed frame-based adaptive CS using OMP reconstruction algorithm.

Fig 2 shows the time domain waveform of the original speech signal and adaptive CS reconstructed speech with different compressed ratio of 0.2, 0.4 and 0.6. As the figures show, based on our proposed frame based adaptive CS framework, a good reconstruction result can be obtained when sufficient number of measurements is given.

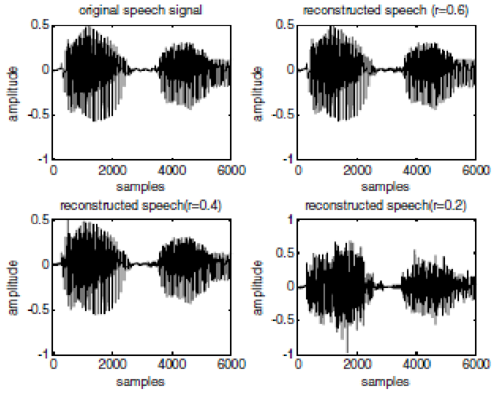

Figure 2. Waveform of original speech and adaptive CS reconstructed speech under different compressed ratio

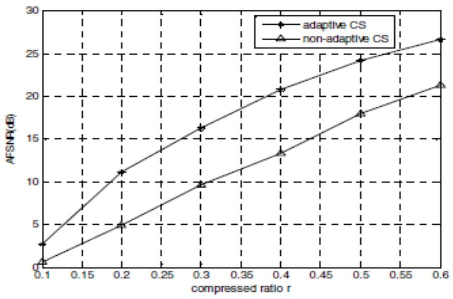

Figure 3. Reconstructed speech quality using the proposed adaptive CS and non-adaptive CS

Fig 3 shows the reconstruction quality of speech signal based on our proposed frame-based adaptive CS and the convential non-adaptive CS. From the fig 3, we can find that reconstruction quality improve when we take adaptive framework. For example, when the compressed ratio is 0.2, the AFSNR increases more than 6dB. Table 1 gives the MOS of the reconstructed speech based on our proposed frame-based adaptive CS and the convential non-adaptive CS. Under each different compressed ratio, we can see that reconstructed speeches using the proposed adaptive frame have higher MOS than that using non-adaptive frame.

TABLE I. MOS of the reconstructed speech using non-adaptive CS and the proposed adaptive CS

compressed ratio	0.1	0.2	0.3	0.4	0.5	0.6
non-adaptive	0.575	1.373	2.084	2.521	3.020	3.476
adaptive	0.813	1.769	2.492	2.935	3.292	3.614

V.CONCULSION

This paper proposes an adaptive frame-based CS framework for speech signals which adjusts the number of measurements of speech frames according to their intra-frame correlations. Our experimental results show that the proposed framework can lead to better CS reconstruction quality than the traditional CS framework. The adaptive compressed sensing that explores the speech signal features to achieve high sampling efficiency creates a new direction for future research on speech signal processing.

VI. REFERENCES

[1] Donoho D L. Compressed sensing [J]. IEEE Transaction on Information Theory, 2006, 52(4):1289- 1306

[2] Donoho D,Tsaig Y. Extensions of compressed sensing [J]. Signal Processing,2006,86(3):533-548.

[3] Scott S. chen,David L. Donoho,Michael A.Saunders. Atomic decomposition by basis pursuit [J]. 2001 Society for Industrial and Applied Mathematics, SIAM review,43(1):129-159.

[4] J. Tropp, A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transaction on Information Theory [J], 2007, 53(12),4655-4666.

[5] Zh. M Wang, G. R. Arcet, J. L. Paredest. Colored random projections for compressed sensing. ICASSP,2007: 873-876

[6] Zhaorui Liu, Vicky Zhao,A. Y. Elezzabi. Block-based adaptive compressed sensing for video. Proceedings of 2010 IEEE 17th international conference on image proceeding,2010,1649-1652.

[7] Jarvis Haupt,Robert Nowak,Rui Castro.Adaptive sensing for sparse signal recovery .ICASSP,2009,702-707.

[8] R.Gribonval, M. Nielsen. Sparse representations in unions of bases. IEEE Trascation on Information Theory, 2004,49(12),3320-3325.

[9] J. A. Tropp. Greed id good:Algorithmic results for sparse approximation. IEEE Trascation on Information Theory,2004,50(10),2231-2242