RESEARCH ARTICLE OPEN ACCESS

Frame-Based Compressed Sensing Of Speech Signal

B. Chaitanya
Assistant Professor, GITAM UNIVERSITY, Hyderabad

Abstract
In this paper, compressed sensing (CS) of speech signal using Frame-based adaptive technique has been proposed. we propose 3 sampling strategies in a frame-based adaptive CS framework. The first stage is frame based adaptive CS framework. In this stage, each speech sequence is divided into non-overlapping frames and all frames in a speech sequence are processed independently. Second stage is Partial Sampling and Frame Analysis. This stage can estimate the amount of intensity changes in the two frames. Last stage is adaptive sampling was used to reconstruction the signal using orthogonal matching pursuit (OMP). In the experimental results shows that the significant improvement of speech reconstruction quality by using frame based adaptive technique compare to the non-adaptive technique.
Keywords-adaptive compressed sensing, speech signal, orthogonal matching pursuit.

I. INTRODUCTION

The key objective in compressed sensing (also referred to as sparse signal recovery or compressive sampling) is to reconstruct a signal accurately and efficiently from a set of few non-adaptive linear measurements. Compressive sensing gives a good solution to overcome the drawback of recover the signal, using the down-sampling method. Candes et al [1] and Donoho [2] proposes that compressive sensing can directly sample linear measurements to recovery the signal without the intermediate stage, hence it reduce the load of sampling the signal but it occur nonconvex problem at reconstruction signal. The approach in [3] is an effort to solve the nonconvex problem in an efficient way. Moreover, Orthogonal Matching Pursuit (OMP) [4] and Tree-Based Orthogonal Matching Pursuit (TOMP) [5] belong to Matching Pursuit algorithm, which are much faster based on the adaptive algorithms.

Most work in CS research focus on random projection matrix which is constructed by considering only the signals' sparsity 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 framework for speech signals, which applies adaptive projection matrix [6].

In this paper is organized as follows. Frame based adaptive CS in section II. Section III describes the OMP. The simulation results are presented in Section IV. Concluding remarks are made in Section V.

II. FRAME BASED ADAPTIVE CS

In this paper, the testing speeches are chosen from CASIA Chinese speech library which is built by the China analysis institute of automation. The testing speech consists of 200 Chinese speech signals and frequency is 16 kHz sampled and 16 bits quantized for each sample. The 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. The flow chart of frame based adaptive CS as shown in fig.1.

a. The Frame-based Adaptive CS framework

The speech signal sparse in 1×n frames and all frames in a speech sequence are processed independently. In this stage, first collect the projections from all frame and compare the projections of current frame and projections of previous frames.

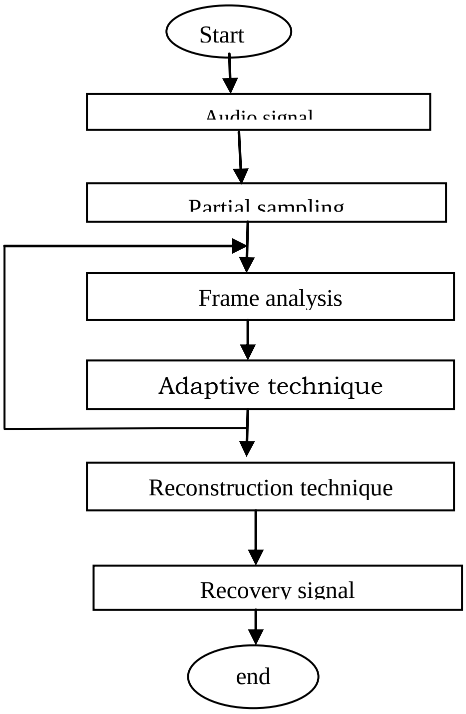

Fig.1. Flow chart of proposed method

Second, we estimate the correlation between previous and current projections and classify the correlation into different categories are explained below.

b. Partial Sampling and Frame Analysis of frame based adaptive CS

The correlation between the two neighbor frames, and can be used classify the correlation. The algorithm of partial sampling and frame analysis as shown below

If $x_t - x_{t-1}$ *= sampling stage*
$Y_t = x_t - x_{t-1}$
$y_d = Y_t$
end
If $x_t - x_{t-1}$ *# sampling stage*
$Y_t - Y_{t-1} = \Phi(x_t - x_{t-1})$
$y_d = Y_t - Y_{t-1}$
end
If $||y_d||_2/Mo <= T_1$
Two neighboring frames are as surd vs. surd.
end
If $T_1 < ||y_d||_2/Mo <= T_2$
Two neighboring frames are sonant vs. sonant.
end
If , $||y_d||_2/Mo > T_1$
Two neighboring frames are surd vs. sonant.
end
where
x_t =current frame
x_{t-1} =previous frame
Y_t=projection matrix of x_t
Y_{t-1}=projection matrix of x_{t-1}
$||y_d||$ norm normalized by Mo and compare with two thresholds T_1 and T_2

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 Mo measurements ytMo collected in the partial sampling stage are sufficient and we do not need additional sampling. That is $y_t = y_t^{Mo}$. 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 M1 (M1>M2) measurements. We use the (Mo+1)th to the M1 th rows of the Gaussian random matrix Φ and combine with y_t^{Mo} to generate the final projection vector yt . The frames that experience large changes must contain the most new information. Therefore, we collect a total of M2 (M1> M2>M3) measurements (including the initial Mo measurements) during the sampling process. The total projection matrix is the first M2 rows of the Gaussian random matrix Φ .

III. ORGTHOGONAL MATCHING PURSUIT (OMP)

In this paper, reconstructing frame by frame of signal using OMP method. OMP uses sub Gaussian measurement matrices to reconstruct sparse signals. If Φ is such a measurement matrix, then Φ∗Φ is in a loose sense close to the identity. Therefore one would expect the largest coordinate of the observation vector y = Φ∗Φx to correspond to a non-zero entry of x. Thus one coordinate for the support of the signal x is estimated. Subtracting off that contribution from the observation vector y and repeating eventually yields the entire support of the signal x. OMP is quite fast, both in theory and in practice, but its guarantees are not as strong as those of Basis Pursuit. The algorithm's simplicity enables a fast runtime. The algorithm iterates s times, and each iteration does a selection through d elements, multiplies by Φ∗ Φ, and solves a least squares problem.

IV. EXPERIMENTAL RESULTS

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 16 kHz 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. 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:

$$AFSNR = \frac{1}{K}\sum_{k=1}^{K} 10\log 10\left(\frac{\left\|x_k\right\|_2^2}{\left\|x_k - \hat{x}_k\right\|_2^2}\right)$$

Where x_k is the orginal speech signal and $\hat{x}_k$ is the reconstruction signal.

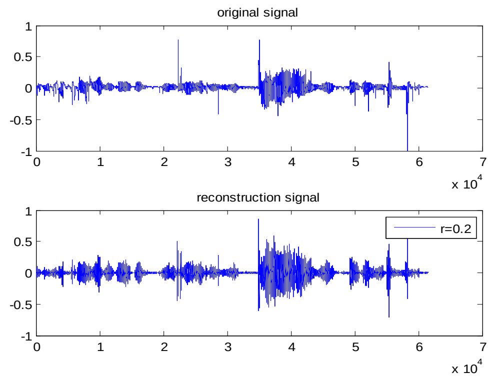

Fig.2. Waveform of original speech and adaptive CS reconstructed speech under r=2

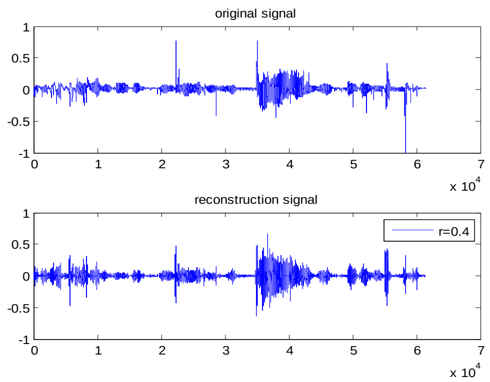

Fig.3. Waveform of original speech and adaptive CS reconstructed speech under r=4

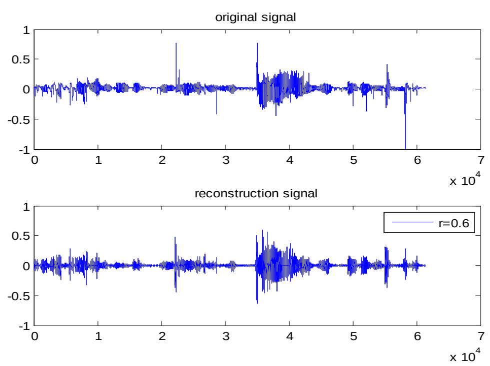

Fig.4. Waveform of original speech and adaptive CS reconstructed speech under r=6

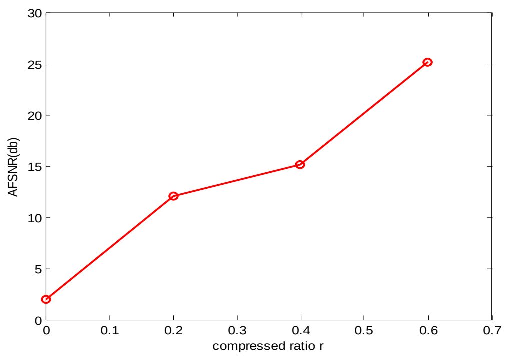

Fig.5.Reconstructed speech quality using the proposed adaptive CS

V. CONCLUSION

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.

REFERENCES

[1] E.J. Candes, J. Romberg, and T. Tao. *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*. Information Theory, IEEE Transactions on, 52(2):489–509,Feb 2006. ISSN 0018-9448.

[2] D.L. Donoho. *Compressed sensing*. Information Theory, IEEE Transactions on, 52(4):1289–1306, April 2006. ISSN 0018-9448.

[3] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, and R.G. Baraniuk. *Distributed compressed sensing of jointly sparse signals*. Signals,Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on, pages 1537–1541, October 28 - November1, 2005. ISSN 1058-6393

[4] Shihao Ji and Lawrence Carin. *Bayesian compressive sensing and projection optimization*. In ICML '07: Proceedings of the 24th international conference on Machine learning, pages 377–384. ACM, New York, NY, USA, 2007. ISBN 978-1-59593-793-3.

[5] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. *Atomic decomposition by basis pursuit*. SIAM Journal on Scientific Computing, 20(1):33–61, 1999.

[6] Qian Xu, Yunyun Ji "*Frame-Based Adaptive Compressed Sensing of Speech Signal*" 978-1-4244-6252-0/11©2011 IEEE