International Journal of Information Processing, 9(2), 1-12, 2015
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

CS Framework for Speech Coding using Sparse Hybrid Dictionary

Vinitha Ramdas[a], Deepak Mishra[a], Sai Subrahmanyam Gorthi[b]

[a]Department of Avionics, Indian Institute of Space Science and Technology, Valiamala, Trivandrum - 695547, India, Contact: vinithar@gmail.com, deepak.mishra@iist.ac.in

[b]Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Trivandrum - 695547, India, Contact: saisubrahmanyam.gorthi@gmail.com

Data compression is immensely important in digital world, with limited resources for storage and transmission. Speech compression or coding deals with efficient digital representation of speech. In this paper, the potential of Compressive Sensing (CS) in the context of speech coding is explored by combining the strengths of dictionary based sparse representations and quantization of CS measurements. CS facilitates effective recovery from fewer measurement samples exploiting the sparsity nature of signals. The major contribution of this work is in designing an easily scalable (medium to high) bit-rate codec by appropriately quantizing CS measurements, where a hybrid dictionary based sparse representation of speech has been used. Most of the earlier CS based speech coding techniques, exploit CS only at the decoder to reduce the processing costs at receiver and compression is achieved by conventional methods. The proposed approach uses CS at the encoding stage itself and provides a performance better than CELP, the widely used medium bit-rate hybrid coding scheme. The work also demonstratres, how effective sparse representations of the speech using a hybrid dictionary can facilitate better compression.

Keywords : AbS Quantization, Compressed Sensing, Hybrid Dictionary, Speech Coding.

1. INTRODUCTION

Speech is an information rich signal which has been the primary means of communication between humans. Digitizing real world signals help to achieve more compact representations and provides better utilization of available resources. With ever increasing demands for system capacity, compression of all real world signals has become a necessity. Researchers have been working on achieving lower bit rates for efficient transmission and storage of digital speech.

Compressed Sensing has gained much attention in literature due to its diverse applications in a variety of fields. Also it exploits the sparsity notion which is a minimally explored, inherent characteristic present in almost all real world signals. CS has been applied to strictly sparse as well as compressible signals. Most of the real world signals are compressible in some domain or the other and this is true for speech signals too. In resource limited scenarios CS can facilitate efficient utilization of the available resources with substantial performance gains. Sparsity is the main principle behind CS and effective sparse representations of signals form the basis for CS based applications.

Conventional hybrid speech coding techniques which transmit at medium bit rates, require parameter extractions and involve linear predictions, pitch updates etc using previously stored samples. This requires huge amount of memory resources. Earlier, most of the CS based speech coding techniques, makes use of CS to reduce dimensionality and processing costs at the receiver side [1][2] (CS does random sampling to obtain M (number of samples acquired) $\ll N$ (original length of the source vector) measurements and the decoder algorithms need to handle only fewer samples). In nary. Figure 6 gives typical speech waveforms for a quantization rate of 2 (worst case results) and a CR of 0.5 (for DCT) and 0.3 (for Hybrid). The waveforms at CR of 0.5 for DCT and CR of 0.3 for hybrid looks similar and resembles original speech indicating that higher compression can be achieved with hybrid dictionary. Thus, the work demonstrates that choice of an appropriate sparsifying transform plays a major role in improving the compression and reconstruction quality of the proposed speech codec.

6. CONCLUSIONS

In this work, the potential of CS in the context of speech coding is explored combining the advantages of dictionary based sparse representation of speech signals and quantization of CS measurements. Scalable bit-rate codecs can easily be designed by using the proposed scheme, just by changing the number of CS measurements that are acquired or by changing the quantization rate per sample. Quantization of CS measurements which facilitated bit rate reduction helps to exploit the advantage of CS (acquiring lesser no: of samples) at the encoder itself. The simulation results demonstrate that the codec can aid transmission at high and medium bit rates. The coder provides higher output SNRs compared to CELP, at the same transmission rates. However, hardware implementation feasibility, delay and complexity of the coder needs to be assessed in detail, for standardization and comparison with other existing ITU standards.

Our study concludes that better sparse representations improve the reconstruction quality of CS systems. The hybrid dictionary used here, provides better performance without imposing much penalty on transmission bit-rates. Moreover, the number of samples acquired, can be reduced to as small as 30% with the hybrid dictionary as compared to 50% for DCT to achieve the same reconstructed speech quality. The hybrid dictionary helped to achieve better compression and higher quality speech at the decoder. In essence, compressed sensing is indeed a suitable candidate for speech coding in resource limited scenarios with advantages of easy scalability and improved efficiency especially when dictionary based sparse representations are used.

REFERENCES

1. Liban A Kassim, Othman O Khalifa and T S Gunawan. Compressive Sensing Based Low Bit Rate Speech Encoder, in *IEEE International Conference on Computer and Communication Engineering (ICCCE)*, pages 302–307, 2012.
2. Gunawan T S, Khalifa O O, Shafie A A and Ambikairajah. Speech Compression using Compressive Sensing on a Multicore System, in *International Conference on Mechatronics (ICOM)*, pages 1–4, 2011.
3. Michelle L Daniels and Bhaskar D Rao. Compressed Sensing based Scalable Speech Coders, in *IEEE Asilomar Conference on Signals, Systems and Computers*, pages 92–96, 2012.
4. Argyrios Zymnis, Stephen Boyd and Emmanuel Candis. Compressed Sensing With Quantized Measurements, in *IEEE Signal Processing Letters*, 17(2):149–152, 2010.
5. Shirazinia A, Chatterjee S and Skoglund M. Analysis-by-Synthesis Quantization for Compressed Sensing Measurements, in *IEEE Transactions on Signal Processing*, 61(22):5789–5800, 2013.
6. Simon Foucart and Holger Rauhut. A Mathematical Introduction to Compressive Sensing, *Springer Science, Business Media, New York*, XVIII, 2013.
7. Andreas S and Spanias. Speech Coding: A Tutorial Review, in *Proceedings of the IEEE*, 82(10), 1994.
8. A M Kondoz. Digital Speech - Coding for Low Bit Rate Communication Systems, *Second Edition, UK*, 2004.
9. Dalei Wu, Wei-Ping Zhu and M N S Swamy. On Sparsity Issues in Compressive Sensing based Speech Enhancement, in *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 285–288, 2012.
10. Maria G Jafari and Mark D Plumbey. An Adaptive Orthogonal Sparsifying Transform for Speech Signals, in *Proceedings of ISCCSP*, pages 786–790, 2008.
11. Maria G Jafari and Mark D Plumbey. Fast Dictionary Learning for sparse representations of Speech Signal, in *IEEE Journal of Selected*

Topics in Signal Processing, 5(5):1025–1031, 2011.
12. Michal Aharon, Michael Elad and Alfred Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, in *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
13. Ron Rubinstein, Alfred M Bruckstein and Michael Elad. Dictionaries for Sparse Representation Modelling, in *Proceedings of the IEEE*, 98(6):1045–1057, 2010.
14. Yue Wang, Zhixing Xu, Gang Li, Liping Chang and Chuanrong Hong. Compressive Sensing Framework for Speech Signal Synthesis Using a Hybrid Dictionary, in *IEEE International Congress on Image and Signal Processing*, pages 2400–2403, 2011.
15. Nowak R D and Wright S J. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems, in *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007.
16. C Sanderson. Biometric Person Recognition: Face, Speech and Fusion, in *VDM-Verlag, ISBN 978-3-639-02769-3*, 2008.

Vinitha Ramdas has been working with ISRO Satellite Centre, Bangalore, India after completing her BTech Degree in Electronics and Communication Engineering from Kerala University since 2006. She completed MTech in Digital Signal Processing from Indian Institute of Space Science and Technology (IIST), Thiruvananthapuram, Kerala, India in 2014. Her research interests include Speech and Image processing, Digital Communications and Pattern Recognition and Machine Learning.

Deepak Mishra is with the Department of Avionics, IIST. He is a post graduate from Devi Ahilya University, Indore (M.P), India and doctorate from Indian Institute of Technology, Kanpur, India. His research interests include Signal and Image processing, Artificial Neural Networks Computer Vision and Graphics and Machine Learning.

Sai Subrahmanyam Gorthi is with the Department of Earth and Space Sciences, IIST. He is a post graduate from Indian Institute of Science, Bangalore, India and doctorate from Indian Institute of Technology, Madras, India. His research interests include Signal and Image processing, Computer Vision and Pattern Recognition.