

A Comparative Study on Compression and Compressed Sensing of Speech Signals

Ms. Flavita Janice Pinto

Student, M.Tech (DECS)
Department of E & C Engineering
St Joseph Engineering College,
Mangaluru, D.K

Ms. Rashmi H

Assistant Professor
Department of E & C Engineering
St Joseph Engineering College,
Mangaluru, D.K

Abstract— Speech processing is the fastest growing technology due to its applications in various fields such as research, forensic and aid for blind people. This paper describes speech processing techniques which involve improving the signal to noise ratio, reducing the compression rate and decreasing the bandwidth required for transmission involving minimum error in the signal at the receiver end.

The speech compression and compressed sensing is done using MP3 technique which is basically compression and decompression using DCT-IDCT technique as well as a comparative study between compressed sensing and compression of speech signals based on Word Error Rate (WER), Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE) and Compression Ratio using MATLAB R2009a as software.

Index Terms—compressed sensing, Discrete Cosine Transform (DCT), Inverse Discrete Cosine Transform (IDCT), Compression.

I. INTRODUCTION

Speech is a form of communication which involves lot of redundancy. Speech requires a lot of storage space as well as large number of bits for transmission. Speech processing is the application of digital signal processing to the processing or analysis of speech signals [1]. The purpose of speech compression is to reduce the number of bits required to represent speech signals. This is done by reducing redundancy in order to minimize the requirement for transmission bandwidth or to reduce the storage cost without affecting the quality of speech at the receiver end.

Compressed Sensing (CS) is an emerging technique that promises to effectively recover a sparse signal from far fewer measurements than its dimension. Compressed Sensing assures almost an exact recovery of a sparse signal if the signal is sensed randomly where the number of the measurements taken is proportional to the sparsity level and a log factor of the signal dimension [8].

Applications of speech processing [2] include speech coding, speech recognition, speech verification, speech enhancement and speech synthesis.

A. Speech signal processing

Speech signal processing is the intentional alteration of auditory signals. There are two types of processors: analog and digital processors. Analog processors operate on electrical signal, while digital processors operate on the

digital representation of analog signal. Analog signal is a mathematical representation of a signal by a set of continuously changing values. Digital representation of a signal is usually in binary form.

II. PROCEDURE

The block diagram of the proposed system is as shown in Figure 1. The entire system is divided into two phases and is carried out in MATLAB. The first stage is the training phase and the second stage is the testing phase. In the first stage i.e., the training phase, the samples of different speakers is collected using the Voice box tool in MATLAB.

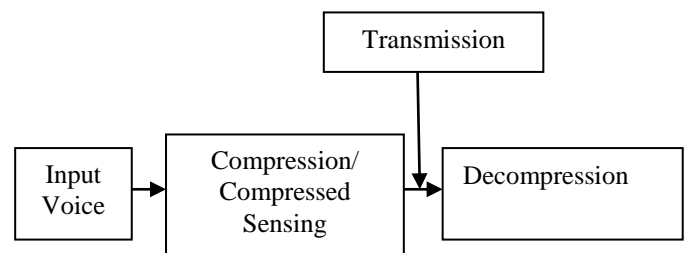


Figure 1. Proposed System

The second stage is the testing phase where a voice is given as input to the system. This voice is compressed using DCT compression technique, transmitted and at the receiving end it is decompressed.

The same procedure is followed for compressed sensing and the two techniques are compared. The comparison is done based on Word Error Rate (WER), Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE) and the compression ratio.

III. BASIC IDEA OF COMPRESSION AND COMPRESSED SENSING

A. Compression

A program for storing the database of voice inputs is written using MATLAB. Using a transducer the speech is given as input. The input is read using the wavread () command and stored in database using wavwrite () command. A .wav format file is taken as input. The sampling rate and the number of samples are calculated. DCT (Discrete

Cosine Transform) is applied on the input signal. The signal and data is compressed. The weighted coefficients are calculated, cut off frequency is specified and the high and low precision values are found for quantization.

Figure 2 represents the flow diagram for the training phase. In this phase the speech is recorded and stored in the database. The sampling frequency and the duration for the initial silence are initialized. The speech is recorded and stored in the database as a .wav file.

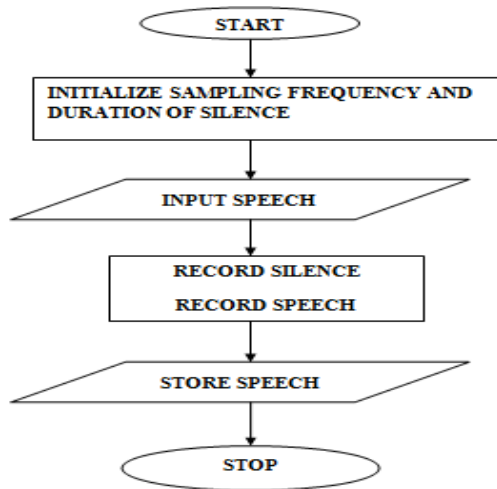


Figure 2. Flow diagram for the training phase

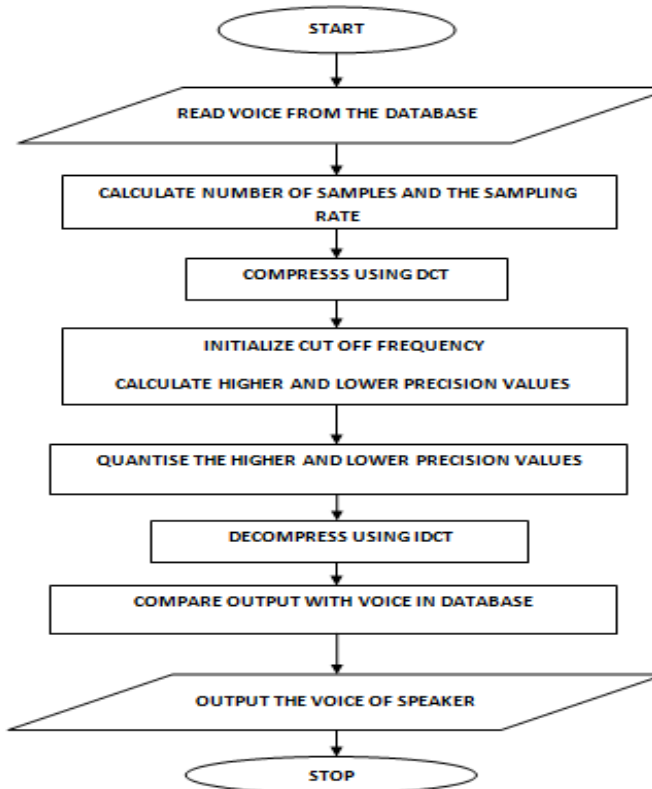


Figure 3. Flow diagram for the testing phase

Figure 3 represents the flow diagram for the testing phase of voice compression-decompression. The voice from the database is taken as input. The sampling rate and the number of samples in the input signal are calculated. The signal is then compressed using DCT (Discrete Cosine Transform). The cut off frequency is initialized in order to calculate the higher and the lower precision values. Using these higher and lower precision values, quantization is performed. The signal is later decompressed and voice recognition is performed.

B. Compressed Sensing

The objective of Compressed Sensing (CS) is to increase the data rates of current and possibly future generation systems. In the proposed system the speech signal is sampled below the Nyquist rate by using compressive sensing. Figure 4 shows the use of compressive sensing in a communication system.

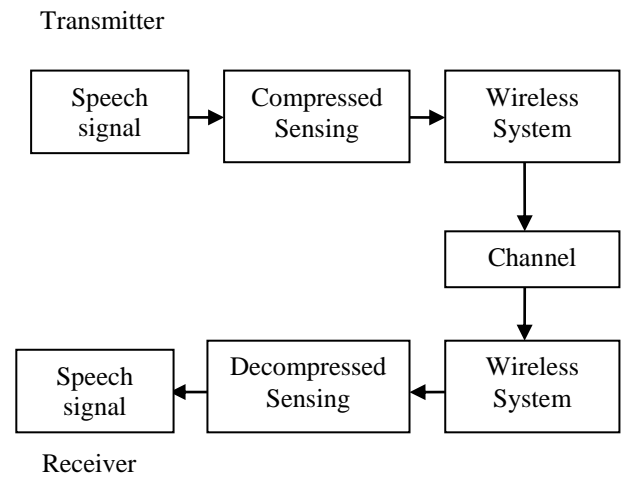


Figure 4. Block diagram for compressed sensing

The compressed spectrum is then transmitted over the wireless system and successfully reconstructed at the receiver without losing any significant information. In the first stage a speech signal is modeled using a Laplace random number generator in MATLAB. It is decided to use a Laplace number generator to model the speech signal, because these types of signals typically have a Laplacian distribution [9]. The modeled speech signal was mapped into the discrete frequency domain using the discrete cosine transform (DCT).

In the second stage, before compressive sensing is applied to the signal, a threshold window is used to eliminate the coefficients that are less significant to the signal. In other words, all the coefficients with small amplitude are multiplied by zero. The purpose of the threshold is to ensure that the DCT spectrum is sparse.

In the third stage, the threshold spectrum is multiplied by the measurement matrix, which is a matrix composed of random numbers. The output of the compressive sensing algorithm is converted into a digital signal using an Analog-to-Digital converter in order to be transmitted by the mobile system. At the receiver section, an initial guess was made using the measurement matrix and the observation vector (vector signal), which is close to the input speech signal. Finally, the speech signal is reconstructed from a significant

small number of observations by using one of the optimization techniques available. The difference between the actual signal and the reconstructed signal is calculated in order to observe the error between both signals.

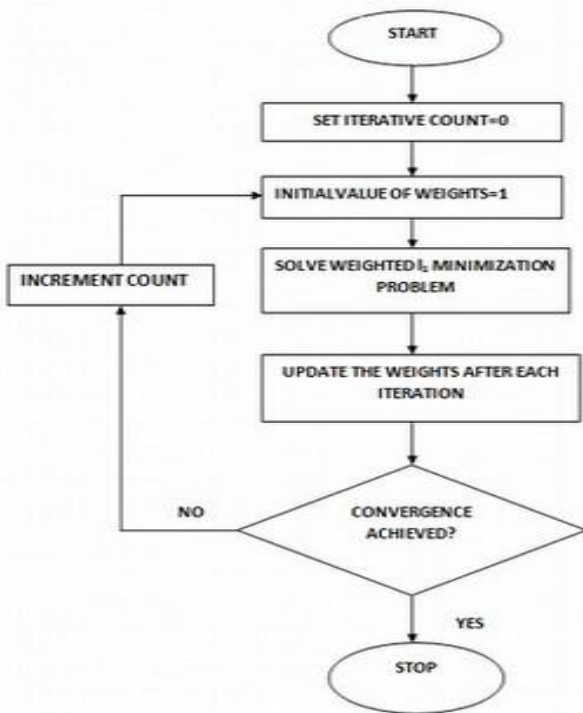


Figure 5. Flow diagram for iteratively reweighted l_1 minimization method for Compressed Sensing [3]

IV. EXPERIMENTAL RESULTS

The simulation results for the three .wav files are shown below. The 1.wav file as shown in Figure 6 consists of a word which is sampled at a rate of 16000 samples per second which results in 32000 samples. DCT is applied to these 32000 samples and the output is a compressed signal as shown in Figure 7. This compressed signal lies in the low frequency region.

A cut-off frequency of 0.00015 is selected. Using this cut-off frequency a mask is applied and higher and lower precision values are calculated. The plot highlighting the higher and lower precision values are shown in Figure 8. Using these high and low precision values the IDCT (Inverse Discrete Cosine Transform) of the samples is plotted as shown in Figure 9.

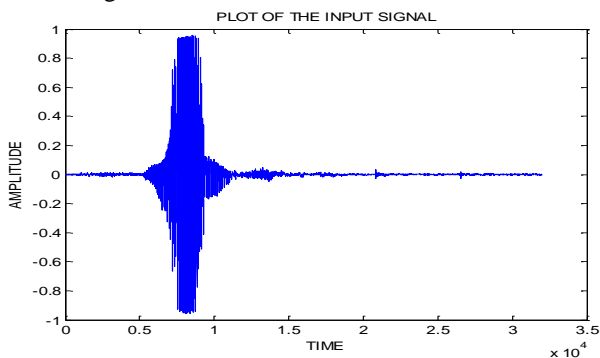


Figure 6. Plot of input signal

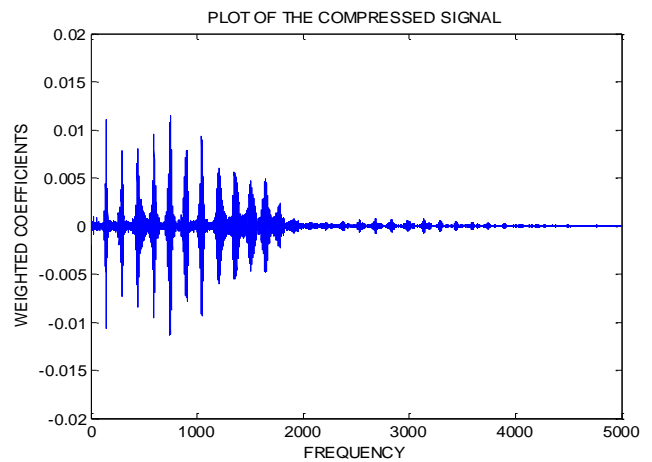


Figure 7. Plot of compressed signal

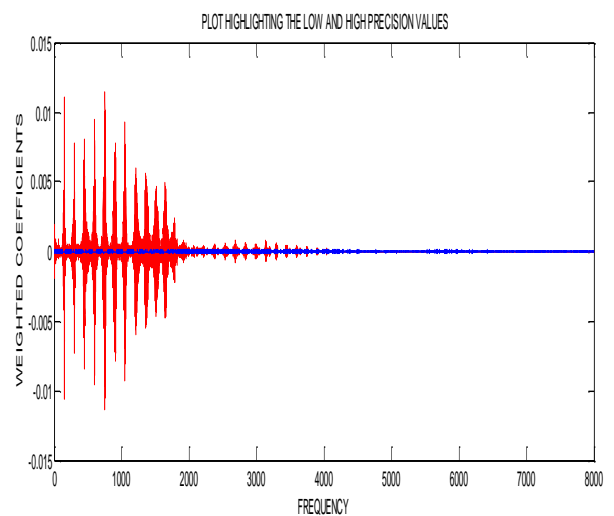


Figure 8. Plot highlighting the low and high precision values

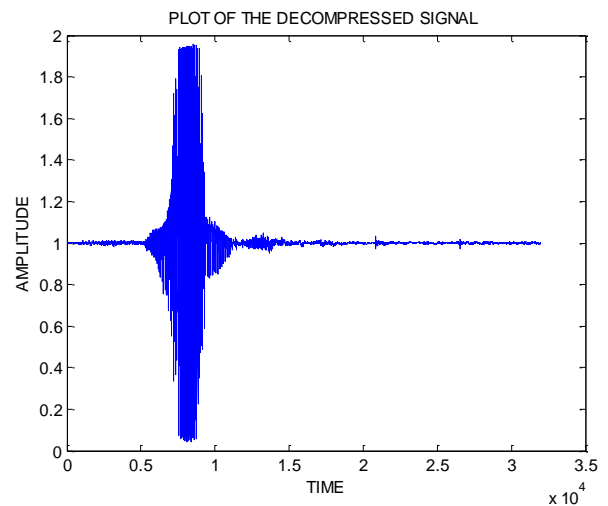


Figure 9. Plot of decompressed signal

Same procedure is followed with different signals and MSE, compression ratios and PSNR are calculated for every signal.

Consider the same input signals which were considered for compression. Compressed sensing is done according to the l_1 minimization technique. The 1.wav signal is taken as input as shown in the Figure 10. DCT is applied to this signal and the signal is compressed as shown in Figure 11.

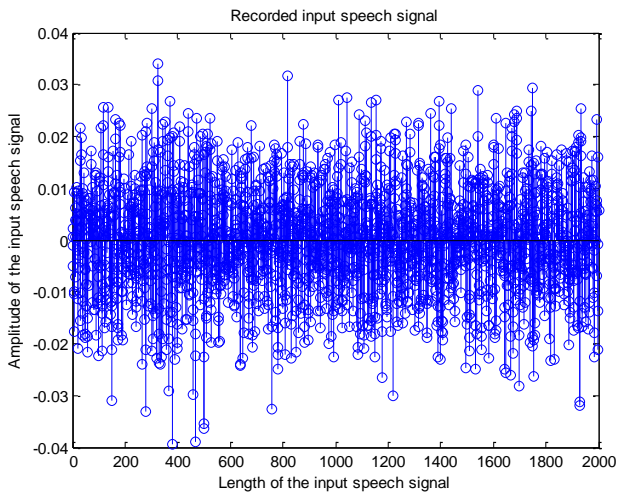


Figure 10. Plot of the input signal

Thresholding of the signal is done to make the signal sparser as shown in Figure 12. This signal is multiplied with a predefined measurement matrix. This results in a vector which is also called as observation vector. Figure 13 shows the reconstructed signal at the receiver.

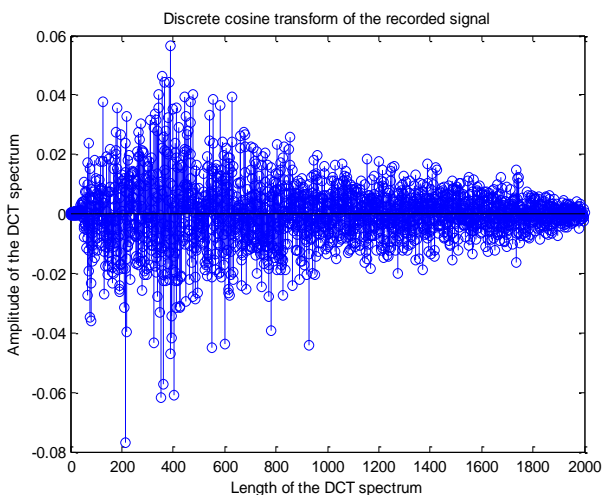


Figure 11. Plot of the DCT

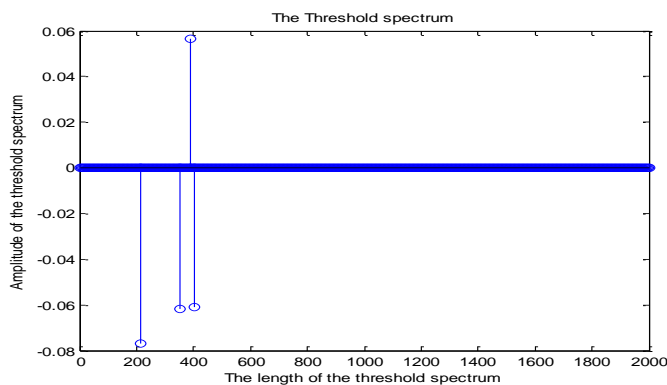


Figure 12. Plot of the threshold spectrum

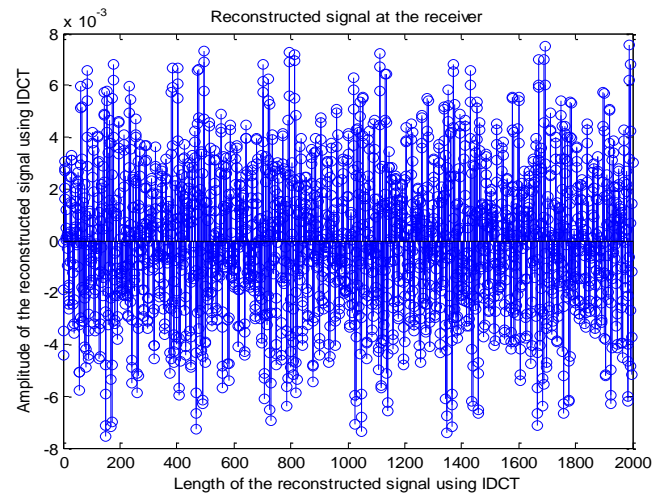


Figure 13. Plot of the reconstructed signal

V. CONCLUSION

From the experimental results it is observed that the Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) and Compression Ratio of Compression and Compressed Sensing are obtained and compared. The following results are obtained as shown in Table 1.

Table 1. COMPARISON

PARAMETERS	COMPARISON	
	COMPRESSION	COMPRESSED SENSING
MSE	MORE	LESS
PSNR	LESS	MORE
COMPRESSION RATIOS	LESS	MORE

From Table 1 we conclude that Compressed Sensing is a better technique when compared to compression for all types of speech signals.

REFERENCES

- [1] en.wikipedia.org/wiki/Speech_processing
- [2] www.ece.ucsb.edu/Faculty/Rabiner/.../341_telecom%20applications.pdf
www.ece.ucsb.edu/Faculty/Rabiner/.../341_telecom%20applications.pdf
- [3] en.wikipedia.org/wiki/Compressed_sensing
- [4] Wei-Ho Tsai, *Member, IEEE*, and Hsin-Chieh Lee “Singer Identification Based on Spoken Daitain Voice Characterization” IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 8, OCTOBER 2012.
- [5] In the paper titled “Door Phone Embedded System for Voice Based User Identification and Verification Platform” published by Iztok Kramberger, *Member, IEEE*, Matej Grašič, and Tomaž Rotovnik IEEE Transactions on Consumer Electronics, Vol. 57, No. 3, August 2011.
- [6] A.A.M. Abushariah, M.A.M. Abushariah, “Voice Based Automatic Person Identification System Using Vector Quantization” International Conference on Computer and Communication Engineering (ICCE 2012), 3-5 July 2012, Kuala Lumpur, Malaysia.

- [7] M. Abdollahi, E. Valavi, H. Ahmadi Noubari "Voice-based Gender Identification via Multiresolution Frame Classification of Spectro-Temporal Maps" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009.
- [8] Siddhi Desai , Prof. Naitik Nakrani "Compressive Sensing in Speech Processing: A Survey Based on Sparsity and Sensing Matrix" International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Issue 12) december 2013.
- [9] David L. Donoho, Member, IEEE "Compressed Sensing IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 52, NO. 4, APRIL 2006.