



On the Design and Performance Evaluation of DWT based Compressed Speech Transmission with Convolutional Coding

Javaid Ahmad
Sheikh
PG Department of
Electronics & IT
University of
Kashmir
Hazratbal Srinagar

Sakeena Akhtar
PG Department of
Electronics & IT
University of
Kashmir
Hazratbal Srinagar

Sahir Majeed
PG Department of
Electronics & IT
University of
Kashmir
Hazratbal Srinagar

Mehboob-ul-
Amin
PG Department of
Electronics & IT
University of
Kashmir
Hazratbal Srinagar

Shabir Ahmad
Parah
PG Department of
Electronics & IT
University of
Kashmir
Hazratbal Srinagar

ABSTRACT

With the recent growth and advancement in multimedia based applications, many new techniques capable of producing good quality of compressed speech have been developed. Thus over the past few years there has been enough work done on compression and enhancement of speech signals. In this paper, the application of Discrete Wavelet Transform for speech compression along with the Convolutional Coding for error detection and correction has been described. A convolutionally encoded 8-DPSK modulated bit stream is transmitted through an AWGN channel. At the decoder the received binary bit stream is demodulated and decoded using Viterbi decoder. The compression is performed using Db10 Wavelets with Hard and Soft Thresholding algorithms. To evaluate the performance of the proposed technique, original and the reconstructed signal at the decoder are compared and various performance parameters in terms of Mean Square Error, Peak signal to Noise Ratio, Retained Signal Energy and Compression Ratio have been calculated. From the results obtained, it is observed that speech compression using Discrete Wavelet Transform along with Convolution Coding, shows better performance in terms of PSNR and MSE (high PSNR and low MSE) as compared to speech compression without convolution Coding. The proposed technique has a lot of scope in wireless communications where bandwidth and Quality of Service (QOS) are two main important factors that are taken into considerations.

General Terms

Speech compression, discrete wavelet transform, convolutional coding, hard thresholding, soft thresholding

Keywords

Speech Compression; Db10 Wavelet; Hard Thresholding; Soft Thresholding; Convolution Coding

1. INTRODUCTION

Speech is the randomly occurring signal and has evolved as a primary means of communication between humans. A typical speech sentence signal consists of two main parts: one carries the speech information, and the other includes silent or noise sections that are between the utterances, without any verbal information. The verbal (informative) part of speech can be further divided into two categories: (a) The voiced speech and (b) unvoiced speech. Voiced speech consists mainly of vowel sounds. It is produced by forcing air through the glottis,

proper adjustment of the tension of the vocal cords results in opening and closing of the cords, and a production of almost periodic pulses of air. These pulses excite the vocal tract. Psychoacoustics experiments show that this part holds most of the information of the speech and thus holds the keys for characterizing a speaker. Unvoiced speech sections are generated by forcing air through a constriction formed at a point in the vocal tract (usually toward the mouth end), thus producing turbulence. Being able to distinguish between the three is very important for speech signal analysis. Being highly random in nature, it seems to be a bit difficult to transmit speech as such over a wireless channel as a lot of bandwidth will be required for its transmission or we cannot store speech as such e.g., for recording purposes. So there comes a need of transmitting or storing speech (in order to reduce storage costs) in its compressed form.

Speech compression is basically to remove redundancy from its adjacent samples. The purpose of speech compression is to reduce the number of bits required to represent speech signals (by reducing redundancy) in order to minimize the requirement for transmission bandwidth (e.g., for voice transmission over mobile channels with limited capacity) or to reduce the storage costs (e.g., for speech recording) [1]. Speech compression, especially at low bit rate explores the nature of human speech production mechanism. When we speak, the air from lungs push through the vocal tract and out of the mouth to produce a sound. For some sounds for example, a voiced sound, or vowel sounds of 'a', 'i' and 'u', the vocal cords vibrate (open and close) at a rate (fundamental frequency or pitch frequency) and the produced speech samples show a quasi-periodic pattern. For other sounds (e.g., certain fricatives as 's' and 'f', and plosives as 'p', 't' and 'k', named as unvoiced sound, the vocal cords do not vibrate and remain open during the sound production. The change of the shape of the vocal tract (in combination of the shape of nose and mouth cavities and the position of the tongue) produces different sound and the change of the shape is relatively slow (e.g., 10–100 ms). This forms the basis for the short-term stationary feature of speech signal used for all frame-based speech coding techniques.

There are a number of speech compression techniques, which are broadly categorized as Waveform-based, Parametric-based, Hybrid coding and Transform coding techniques. As the name implied, waveform based speech compression is mainly to remove redundancy in the speech waveform and to

reconstruct the speech waveform at the decoder side as closely as possible to the original speech waveform. Waveform-based speech compression techniques (such as PCM, ADPCM) are simple and normally low in implementation complexity, whereas their compression ratios are also low. Parametric-based speech coding is based on the principles of how speech is produced. It is based on the features that speech signal is stationary or the shape of the vocal tract is stable in short period of time (e.g., 20 ms). The quality of parametric based speech codecs (such as LPC [2-3]) is low, with mechanic sound, but with reasonable intelligibility. Hybrid coding techniques (such as CEPL) were proposed to combine the features of both waveform-based and parametric-based coding. It keeps the nature of parametric coding which includes vocal tract filter and pitch period analysis, and voiced/unvoiced decision. Instead of using an impulse period train to represent the excitation signal for voiced speech segment, it uses waveform-like excitation signal for voiced, unvoiced or transition speech segments. Extending the resolution of the frequency domain decomposition leads to transform coding, wherein a transform is performed on a frame of input speech and the resulting transform coefficients are quantized and transmitted to reconstruct the speech from the inverse transform. Typical examples of transform coding techniques are DCT, DWT, and DWPT etc. [4].

2. LITERATURE SURVEY

Manas Arora et al. (2014) analyzed speech compression with varying bit rate to remove errors and noisy signals which is suitable for remote broadcast lines, studio links, satellite transmission of high quality audio and voice over internet. This paper focuses on speech compression process and its analysis by which processed speech signal can be heard with clarity and in noiseless mode at the receiver end [5].

Sarika R. Gorantiwar et al. (2014) presented a review paper on various speech coders and highlighted their parameters, merits, demerits, types and applications. These coders are capable of producing high quality speech reconstruction at low bit rates. Most of these coders utilize the spectral properties of speech and hence achieve improved compression as compared to direct quantization [6].

D. Ambika and V. Radha (2012) presented a paper on speech compression using Linear Predictive Coding and Discrete Wavelet Transform. Nine samples of speech signal were evaluated and compressed using both LPC and DWT. It was observed that with DWT technique, performance was better for these speech samples than the LPC method [7].

Satish Kumar et-al. (2012) presented a paper on Wavelet Compression Techniques. It is being observed that with Wavelet Compression Technique the ratio of compression and signal to noise of a speech signal can be well-adjusted. ‘Haar’ and ‘db6’ wavelet are used to decompose a signal at level 6 [8].

The above referred techniques [5, 6, 7, and 8] suffer from a number of drawbacks. Technique [5] has a drawback that it cannot be utilized in real time applications because of the fact that no practical approach has been utilized and nothing has been discussed about the transmission of the compressed speech. In our proposed work both the practical implementation and transmission of the compressed speech has been realized. Technique [5] again did not use the pass-band modulation which is pre-requisite in wireless

communication particularly in cellular communications. In our paper an attempt has been made to overcome this drawback. Technique [5 and 6] have used DWT technique for speech compression but no coding or encoding technique has been used. In our proposed work an attempt has been made to overcome all these drawbacks to some extent so that a good quality of coded compressed transmitted speech with low bit rate for cellular environment is obtained.

3. WAVELET TRANSFORM

Wavelet can be defined as a function that does not last forever or a function that has an average value zero. A wavelet is thus a short wave that does not last forever i.e., they are significant in certain period of time and insignificant outside. Wavelet is thus different from sinusoids as a sine wave lasts forever [9]. Figure 1 shows a basic representation of wavelet.

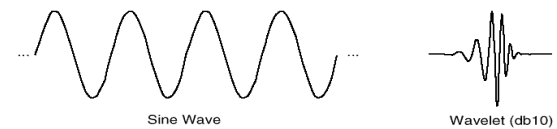


Fig 1: Basic representation of a db10 wavelet

Wavelets have found an important application in speech and image applications. With wavelets a speech/audio signal can be analyzed both in time as well as frequency domain simultaneously. For example when we listen to an audio signal, it does not last forever. We stop after some time. During the time we listen to an audio file, we actually want to understand which frequency contents are being played or what frequency components are predominant during particular time intervals thus it becomes possible with wavelets to analyze a signal both in frequency and time domain. Also with wavelet compression algorithms it is desirable to get back the original information with minimal loss [10].

3.1 Wavelet Transform: Choice of Wavelet

Wavelet is a new technique for examining and comparing a speech signal, it is more advantageous technique because it holds both time and frequency aspect of a signal. Wavelet breaks speech signal into different coefficients. Some of the coefficients having small value are treated as insignificant during data compression and are hence discarded [11]. Wavelets are obtained by a single Mother Wavelet by delay and shifting.

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \quad (1)$$

Where ‘a’ and ‘b’ are called as scaling and shifting parameters. In a high quality speech coder, choosing a Mother wavelet is of key importance. DWT due to its orthogonal properties do not produce any redundant information. The scaling and shifting parameters are thus responsible for Multi-Resolution Analysis algorithm, which decomposes a signal into scales having different time and frequency resolutions. Various Mother wavelet (e.g. Haar, Daubechies, Coiflets, Symlet, Biorthogonal and etc.) functions are differentiated upon these scaling functions, thus the choice of wavelet decides the final waveform shape [12]. In our proposed work Daubechies wavelet ‘Haar’ and ‘db10’ has been used. The scaling filters associated with ‘Haar’ and ‘db10’ are shown in Figures 2 and 3 respectively.

The use of wavelet transform for removing common noise from a speech signal or simply denoising a speech signal has been found much effective as compared to the Fourier Transform. In Telecommunication noise is the main element that limit the capacity of the systems and also in Signal Measurement Systems, the accuracy of the results is also affected by the noise. Thus removing noise from a signal is of core importance mainly in Communication and Signal Processing Systems [13].

In our proposed work db10 wavelet with 10 vanishing moments has been used. The scaling function, analysis and the synthesis filter used in Db10 wavelets are shown in figures 2 and 3 respectively.

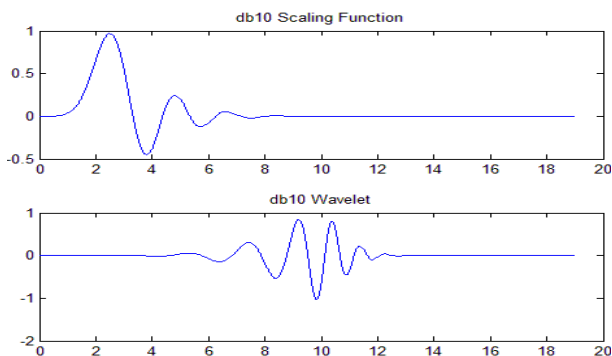


Fig 2: db10 Scaling function and Wavelet

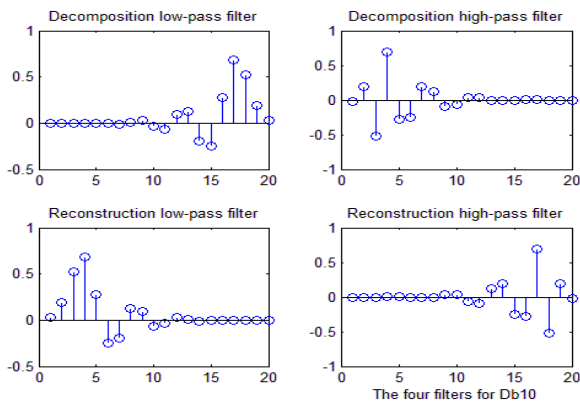


Fig 3: Four Filters of db10 Wavelet

4 CONVOLUTIONAL CODING

A convolutional encoder accepts a sequence of message symbols and produces a sequence of code symbols. Its computations depend not only on the current set of input symbols but, on some of the previous input symbols as well. In practice, convolutional codes operate on a block at a time and so like block codes, have intra-block memory, and possibly no inter-block memory [14]. Convolutional Encoder: A convolutional encoder is made of a fixed number of shift registers. Each input bit enters a shift register and the output of the encoder is derived by combining the bits in the shift register. The number of output bits depends on the number of modulo 2-adders used with the shift registers.

Encoder Parameters: Convolutional codes are commonly specified by the three parameters (n, k, and m) where,

n = number of output bits

k = number of input bits and,

m = number of memory registers.

The quantity k/n called the code rate is a measure of the bandwidth efficiency of the code. Commonly k and n parameters range from 1 to 8, m from 2 to 10, and the code rate from 1/8 to 7/8 except for deep space applications where code rates as low as 1/100 or even longer can be employed. Convolutional codes can also be specified by the parameters (n, k, L) where, L is known as the constraint length of the code and is defined as the number of bits in the encoder memory that affects the generation of the n output

4.1 Viterbi Decoder

The Viterbi decoder algorithm proposed in 1967 is a decoding process for convolutional codes. Convolutional coding, as we all known, has been widely used in communication systems including deep space communications and wireless communications, such as IEEE 802.11a/g, WiMax, DAB/DVB, WCDMA and GSM. Viterbi decoding algorithm is mostly applied to convolutional encoder and it uses maximum likelihood decoding technique [15]. Noisy channels cause bit errors at receiver. Viterbi algorithm estimates actual bit sequence using trellis diagram. Commonly, its decoding algorithm is used in two different forms. This difference results from the receiving form of the bits in the receiver [16]. Decoded information is received with hard decision or soft decision. Decoded information is explained with ± 1 on hard decision operation while soft decision decoding uses multi bit quantization [15].

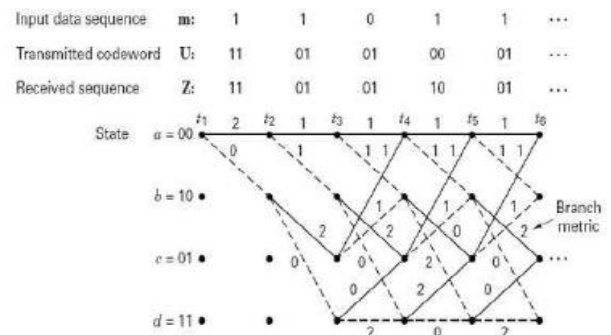


Fig 4: Trellis diagram of Viterbi decoder

Hard decision and soft decision Viterbi decoding refer to the type of quantization used on the received bits. Hard decision decoding uses 1 bit quantization on the received channel values while soft decision decoding uses multi bit quantization on the received channel values [14]. For hard decision decoding, the symbols are quantized to one bit precision while for soft decision decoding, data bits are quantized to three or four bits of precision. The selection of quantization levels is an important design decision because of its significant effect on the performance of the link [17].

5 PROPOSED TECHNIQUE

The proposed technique of speech compression using DWT with convolutional Coding is shown in the figure 5 below. A test recorded speech signal is first decomposed by analysis filter bank of DWT with decomposition level 5. The decomposed speech signal is compressed with db10 wavelet of Daubechies Wavelet Family with both hard and soft thresholding algorithms. The compressed signal is then encoded with convolutional encoder and is modulated with Quadrature Phase Shift Keying Modulation (QPSK)

technique. The modulated signal is sent over Additive White Gaussian Noise Channel for transmission over the network. At the receiving end, the received signal is first demodulated with QPSK demodulator and then decoded with Viterbi decoder. We then try to reconstruct the speech signal from the encoded signal using wavelet reconstruction and finally we convert this reconstructed signal into a speech signal.

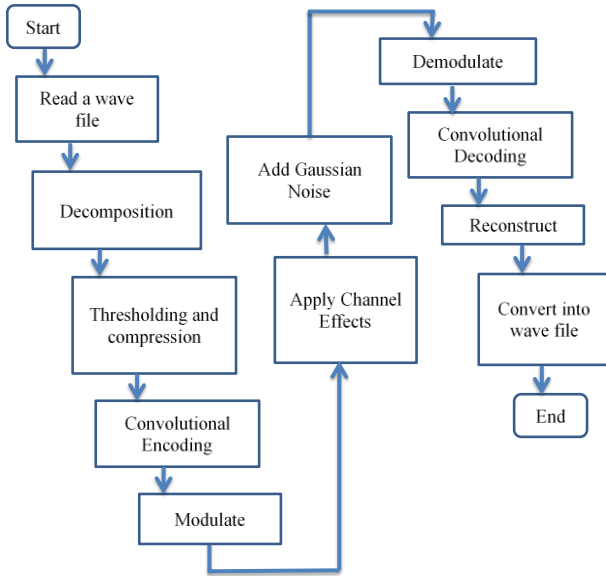


Fig 5: Block Diagram of the Proposed Technique

6 SIMULATION RESULTS

A speech signal is compressed with DWT with and without convolutions coding. We compare the performance of both the techniques i.e., a speech signal compressed using DWT without convolutional Coding and a speech signal compressed using DWT with convolutional coding. The original and reconstructed speech signals are compared and various performance parameters have been calculated in terms of Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), % of Zero Coefficients (PERF0), Retained Signal Energy (PERF2) and Compression Ratio (CR). Table 1 and 2 gives the performance parameters of the proposed techniques. Figure 6 shows the comparative performance and figures 7 and 8 shows the original and reconstructed speech signals with proposed techniques respectively.

Table 1: Performance Parameters of db10 compressed speech signal without Convolutional coding

DWT without Convolutional Coding		
S. No	Hard Thresholding	Soft Thresholding
MSE	0.0053	0.0047
PSNR	70.9247	71.3867
PERF0	95.0208	95.0208
PERF2	97.5721	96.5555
CR	0.9999	0.9999

Table 2: Performance Parameters of db10 compressed speech signal with Convolutional coding

DWT with Convolutional Coding		
S. No	Hard Thresholding	Soft Thresholding
MSE	1.9366-04	3.9312e-04
PSNR	85.2603	82.1855
PERF0	96.5502	96.5562
PERF2	98.1024	87.3875
CR	1	1

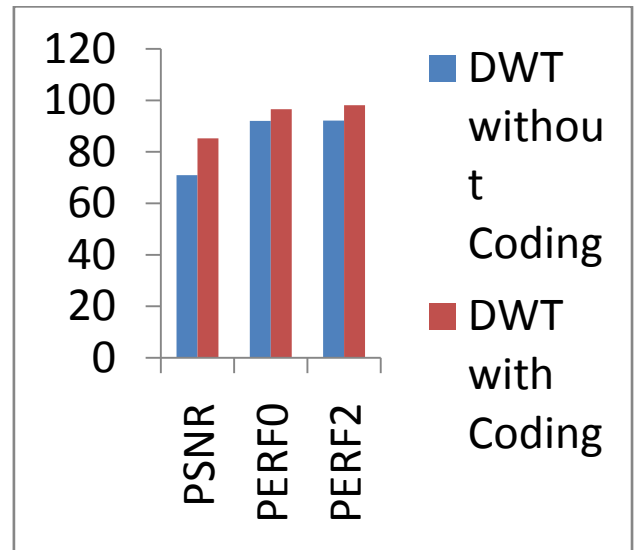


Fig 6: Performance comparison of speech signal compressed using db10 wavelet with and without convolutional coding

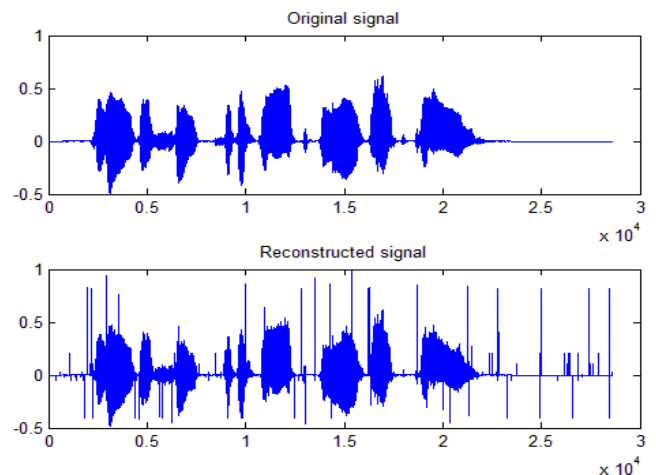


Fig 7: Original and Reconstructed speech signals compressed using db10 without convolutional coding

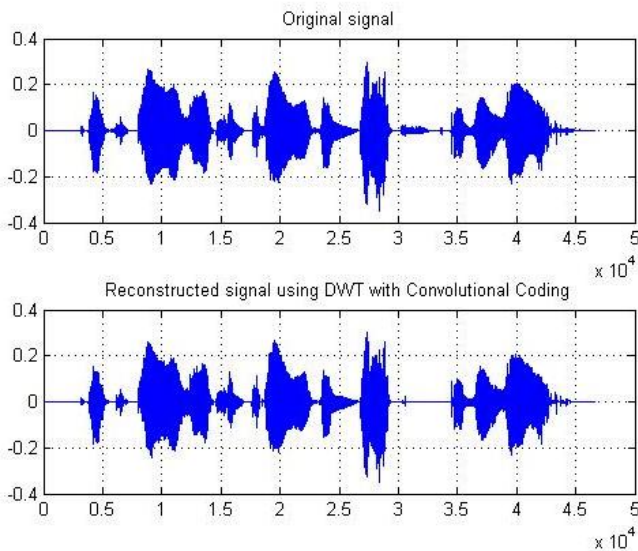


Fig 8: Original and Reconstructed speech signals compressed using db10 with convolutional coding

7 CONCLUSION

This paper presents a comparative study of two compression techniques for compression a test recorded speech signal. In our first attempt, a test recorded speech signal is compressed using db10 wavelet (hard and soft thresholding) without convolutional coding and then, the same test recorded speech signal is again compressed using db10 wavelet with convolutional coding. After successful implementation of both these techniques, various results have been obtained and it was found that among these two techniques, speech compression using db10 with convolutional coding shows better performance in terms of PSNR, Retained Signal Energy and % of zero coefficients. Moreover among the two thresholding algorithms used in this paper, DWT with convolutional coding using Hard thresholding gives higher PSNR value than the soft thresholding.

8 REFERENCES

- [1] Rabiner L.R, and Schafer R.W. Digital Processing Of Speech Signal, Prentice Hall. 2012
- [2] Amol. R. Madane, Zalak Shah, Raina Shah, Sanket Thakur. Speech Compression Using Linear predictive coding, Proceedings of the international Workshop on Machine Intelligence Research, 2009.
- [3] Shiyo. M. Joseph, Firoz Shah A., Babu Auto. Comparing Speech Compression using Waveform coding and parametric coding, International Journal of Electronics Engineering, 3(1), 2011.
- [4] Hemant Amhia and Ratish Kumar. A new approach of speech compression by using DWT and DCT, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.3, July 2014.
- [5] Manas Arora, Neha Maurya, Poonam Pathak and Vartika Singh, “ speech compression analysis using matlab”, International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308, Volume: 03 Issue: 01 | Jan-2014.
- [6] Sarika R. Gorantiwar and Naresh P. Jawarkar, “Speech Coding Techniques: A Review”, IJPRET, Volume 2 (8): 324-330, 2014.
- [7] D. Ambika and V. Radha, “A Comparative Study between Discrete Wavelet Transform and Linear Predictive Coding”, World Congress on Information and Communication Technologies, 978-1-4673-4805-8/12/\$31.00_c 2012 IEEE.
- [8] Satish Kumar, O.P. Singh, G.R. Mishra, Saurab Kumar Mishra, Akanksha Trivedi, “Speech Compression and Enhancement using Wavelet Coders”, International Journal of Electronics Communication and Computer Engineering Volume 3, Issue 6, ISSN (Online): 2249-071X, ISSN (Print): 2278-4209, 2012.
- [9] Gilbert Strang and Truong Nguen, “Wavelets and Filter Banks”. Wellesley-Cambridge Press, MA, 1997, pp. 174-220,365-382
- [10] Andrew K. Chan and Jaideva C. Goswami, “Fundamentals of Wavelets”, Wiley-India Edition, John Wiley & Sons Inc, New Delhi, 1999, pp. 89-97
- [11] Nikhil Rao, “Speech Compression Using Wavelets”, ELEC 4801 THESIS PROJECT. School of Information Technology and Electrical Engineering, the University of Queensland, October 2001
- [12] Brian Gamul kiewicz and Michael Weeks, “Wavelets based Speech Recognition”, Proc. IEEE International Symposium on Micro-Nano Mechatronics of Human Science, Dec.2003, pp. 678-681 Vol. 2, doi: 10.1109/MWSCAS.2003.1562377.
- [13] Ingrid Daubechies, “Ten Lectures on Wavelets”, SIAM, 1992, pp. 115- 132,194-292,258-259
- [14] T. Mc Dermott, “Wireless Digital Communications: Design and Theory”, Tuscon Amateur Packet Radio Corporation, Tuscon, Arizona, 1996.
- [15] J. G. Proakis, “Digital Communications”, 3rd edition, WCB/McGraw-Hill, boston, Massachusetts, 1995.
- [16] John. M. Geist, “Viterbi Decoder Performance in Gaussian Noise and Periodic Erasure Bursts ”IEEE Transactions on Communications, vol. com-28, no. 8, pp.1417-1422, August 1980.
- [17] A.M. Michelson and A. H. Levesque, Error Control Techniques for Digital Communication, John Wiley & Sons, New York, 1985