A Comparative Study of Speech Compression using Different Transform Techniques

Jithin James

Department of Electronics and Communication
Vimal Jyothi Engineering college, Chemperi
Kannur, India

Vinod J Thomas

Department of Electronics and Communication
Vimal Jyothi Engineering college, Chemperi
Kannur, India

ABSTRACT

Compression of speech signal is an important field in digital signal processing. Speech signal compression has significant importance in today's world, because of limited bandwidth and transmission or storage capacity. Speech compression is a process of converting human speech signals into efficient encoded representations that can be decoded back to produce a close approximation of the original signal. This paper explores a transform based methodology for compression of the speech signal. In this methodology, different transforms such as FFT, DCT and DWT are exploited. A comparative study of performance of different transforms is made in terms of SNR, PSNR NRMSE and compression factor (CF). When compared, DWT gives higher compression with respect to DCT and FFT in terms of CF.

Keywords
Speech compression, FFT, DCT, DWT Compression ratio

1. INTRODUCTION
Speech can be defined as the response of vocal tract to one or more excitation signal having bandwidth of 0-4 KHz. Speech coding has been and still is a major issue in the area of digital speech processing. The motivation behind this fact that access to unlimited amount of bandwidth is not possible. Therefore, speech signal should be coded and compressed. In speech compression, it is desired to represent signal by as small as possible number of coefficients within an acceptable loss of quality.

Speech compression techniques are mainly focused on removing short term correlation (in the order of 1ms) among speech samples and long term correlation (in the order of 5 to 10ms) among repeated pitch patterns. Higher the speech bit rate (Kb/s), higher the speech quality and greater the bandwidth and storage requirements [1]. In practice there is always a trade-off between bandwidth utilization and speech quality. Speech compression is extensively used in many applications such as cellular telephony, video teleconferencing systems and mobile satellite communications.

In the last two decades, substantial progress has been made in the field of speech compression. Waveform based speech compression techniques are simple and normally low implementation complexity, where as their CF are low (eg. PCM, ADPCM). In parametric speech coding, a preprocessor is used to extract some features that can be later used to extract the original signal. They have higher implementation complexity, but can achieve better CF. The quality of parametric based speech codecs is low [2] (eg. LPC,CELP).

In past, researches have made in the many transform methods such as FFT, DCT and DWT which extensively used. FFT is a DFT algorithm which reduces the number of computations needed for N points from $2N^2$ to $2Nlog_2N$ [3]. Both one dimensional DCT (DCT-1) two dimensional DCT (DCT-2) are used for data compression [4]. For speech compression, DCT-1 is exploited. DCT can be used for speech compression, because of high correlation in adjacent coefficients. We can reconstruct signal accurately from very few DCT coefficient.

In the last decade, wavelet transform and more specifically DWT has emerged as powerful and robust tool for analyzing and extracting information from speech signal [5]. Recently several other methods have been developed based on wavelets or wavelet packets for compressing speech signal [6]. In this paper, we mainly discussed narrow band speech compression having speech spectrum from 0 to 4 KHz. In recent years, VOIP and mobile applications use wide band (WB) speech (0-7KHz) to provide high fidelity speech transmission quality. ITU-T G.722.1 [7] is an example of WB speech codec using transform coding.

The paper is organized as follows. A brief introduction on speech compression techniques is presented in section-1. Section-2 discusses various transform methods. Section-3 describe compression methodology based on these transforms. Finally section-4 presents performance comparison followed by conclusion.

2. OVERVIEW OF TRANSFORM TECHNIQUES
The three important features of a suitable transform are its compressional efficiency, which relates to concentrating the energy at low frequencies, ease of computation and minimum mean square error. The ideal transform for achieving these features is the karhunen-love transform, but this cannot be represented algorithmically. The following three transforming techniques namely FFT, DCT and DWT are employed for speech compression.

2.1 Fast Fourier Transform
In the development of the DFT, we represented finite length sequences by first forming periodic sequences from which the finite length sequence can be uniquely recovered and then utilizing an expansion in terms of periodic complex exponentials. That is signal can be analyzed or synthesized as a number of harmonically related sine and cosine signals. The fast fourier transform (FFT) is an efficient computation algorithm of the DFT.

Basically, the computational problem for DFT is to compute the sequence X(k) of N complex valued numbers, given another sequence of data x(n) of length N, according to the formula

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{-kN} \quad 0 \leq n \leq \text{N-1}$$

where $W_N = e^{-J2\pi/N}$

similarly the DFT becomes

$$x(n) = \frac{1}{N} \sum_{K=0}^{N-1} X(k) W_N^{-kn} \quad 0 \le n \le N-1$$

We observe that for N numbers of points, direct computation require N^2 complex multiplications whereas FFT algorithm require only $\frac{N}{2}\log_2 N$ complex multiplications.

2.2 Discrete cosine transform

DCT forming a periodic, symmetric sequences from a finite length sequence in such a way that the original finite length sequence can be uniquely recovered. It consists essentially of the real part of the DFT. This definition is reasonable, since the fourier series of a real and even function contains only the cosine terms. There are many ways to do this, so there are many definitions of the DCT.

DCT-1 is used in signal compression applications in preference to the FFT because of a property energy compaction. The DCT-1 of a finite length sequence often has its coefficients more highly concentrate at low indices than the DFT does [8].

The DCT-1 is defined by the transform pair

$$X(k) = \sqrt{\frac{2}{N}} C_k \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi k(2n+1)}{2N}\right) \quad 0 \le k \le \text{N-1}$$

IDCT is defined as

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} C_k X(k) \cos\left(\frac{\pi k(2n+1)}{2N}\right) \quad 0 \le n \le \text{N-1}$$

where weighting function $C_m = \sqrt{\frac{1}{2}} \quad ; \quad k = 0$

$$1 \quad ; \quad 1 \le k \le \text{N-1}$$

2.3 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a special case of wavelet transform that provide a compact representation of a signal in time and frequency. DWT decomposes the signal into mutually orthogonal set of wavelets [9] which is the main difference from CWT. So we get no redundant information and DWT is very good for signal processing and compression.

Basic principle of wavelet transform is that it decomposes the given signal in too many function by using property of translation and dilation of a single function called a mother wavelet is defined as

$$\varphi_{s,\tau} = \frac{1}{\sqrt{s}} \varphi\left(\frac{t-\tau}{s}\right)$$

where 's' is the variable scaling constant and 'τ' is the constant of translation.

The discrete wavelet transform of signal s(k) may be defined as

$$\text{DWT(m,n)} = 2^{-m/2} \sum_k s(k) \varphi(2^{-m}k - n)$$

Thus DWT is obtained by discretizing the scale parameter 's' and the translation parameter 'τ' and setting $s_0 = 2$ and $\tau_0 = 1$ such that $s = s_0^m$ and $\tau = n\tau_0 s_0^m$ where m & n are integers.

2.3.1 Wavelet Decomposition

The choice of the mother wavelet function used in designing high quality speech coders is of prime importance. Several factors are evolving for selecting an optimal wavelet function. Reconstructed signal have minimum reconstructed error variance and maximum signal to noise ratio [10].

Wavelet with more vanishing moments provide better reconstruction quality, as they introduce less distortion into the processed speech and concentrate more signal energy in a few neighboring coefficients. But problem is that the computational complexity of the DWT increases with number of vanishing moments and hence for real time applications it is not practical to use wavelets with an arbitrarily high number of vanishing moments.

2.3.2 Multiresolution Analysis

A multiresolution analysis (MRA) or multi scale approximation (MSA) is the design of DWT and the justification for the algorithm of the fast wavelet transform (FWT). FWT is an efficient algorithm [11] for computing wavelet coefficient by using techniques from the theory of filter banks .

A filter bank [12] is a set of filters, which split up the signal's frequency components into different signals, each with a subset of frequencies.The combined pass bands of the filter cover the entire frequency range, so the filters are complimentary. A simple filter bank consists of one low pass filter and one high pass filter, both having a cutoff frequency at half the frequency bandwidth. Filter design is particularly important in achieving good performance in MRA. It is clear that we cannot use brick wall filter characteristics, since such filters are physically unrealizable. A particularly practical solution to the aliasing problem is to use quadrature mirror filters (QMF), which have the frequency response characteristics shown in fig 1.
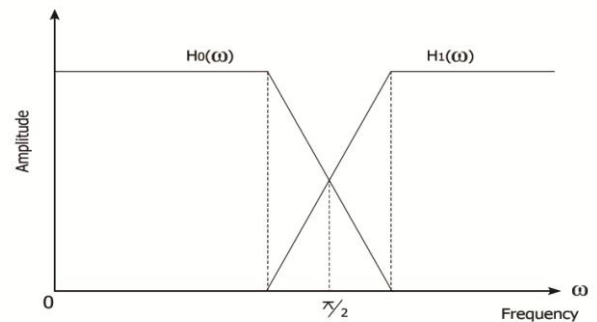

Fig 1: Overlapping of LPF & HPF Filter Responses

To construct a filter bank with more than two frequency bands, y_0 could be filtered again by two filters, one low pass filter and one high pass filter which divide the bands again into two bands. The length of output signal (number of samples) have doubled. The solution is to down sample the signal (or decimation).

For reconstruction, up sampling (or expanding) must be done in order to undo the decimation. Fig 2 shows a two channel filter bank analysis followed by re-synthesis.
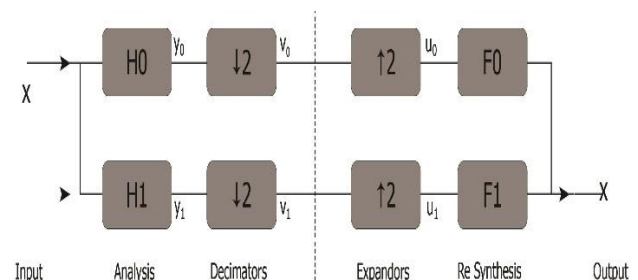

Fig 2: Two Channel Analysis & Re-synthesis Filter Bank

3. COMPRESSION METHODOLOGY

In this paper, we exploits the compression of speech signal using transform method [13]. Transform techniques do not compress the signal, they provide information about the signal and using various encoding schemes, compression of signal is done. Compression is achieved by neglecting small magnitude coefficients as insignificant data and thus discarding them.

The process of compression of speech is carried out by using number of stages as shown in fig 3. In first stage, that is in decomposition step, input speech signal is decomposed into different resolutions or frequency bands. After decomposition, compression involves truncation and quantization of coefficients followed by encoding and resulting compressed speech signal.
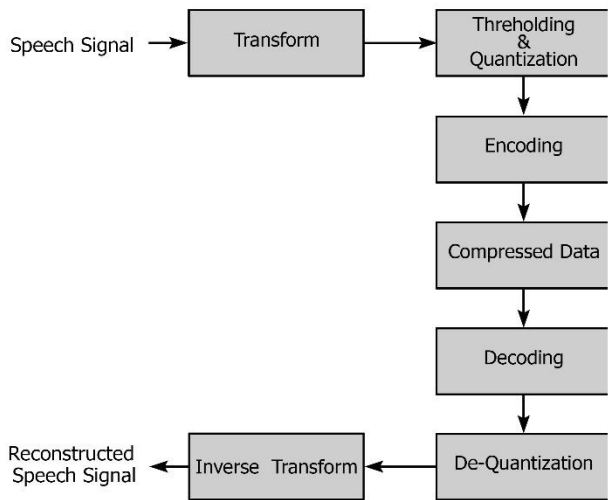

Fig 3: Design Flow of Speech Coder

3.1 Thresholding

After calculating the transform coefficients of the speech signal, compression involves truncating the transform coefficients below a threshold. Most of the coefficients, more than 90% of transform coefficients were found to be insignificant and truncated to zero. This means that most of the speech energy is in high valued coefficients which are few. There are two different approaches are available for calculating threshold [14].

3.1.1 Global Threshold

Global thresholding involves thresholding every sub band (sub signal) with the same threshold value. The global threshold values are chosen from the range $(0,\text{---}c_{max})$, where c_{max} is the maximum coefficient in the decomposition.

3.1.2 Level Dependent Threshold

This approach consists of applying visually determined level dependent thresholds to each decomposition level. The truncation of insignificant coefficients can be optimized when such a level dependent thresholding is used. Thus this method of thresholding selects the highest absolute valued coefficients at each level.

3.2 Quantization

Quantization is a nonlinear and irreversible operation that maps continuously valued input data to set of discrete valued output data. The aim of quantization is to decrease the information found in the transform coefficients in such a way that the process brings perceptually no error. There are two types of quantization are available.

3.2.1 Uniform Quantization

For uniform quantization, the step size Δ is given by

$$\Delta = \frac{m_{max} - m_{min}}{L}$$

where m_{max} is the maximum value in the signal matrix, m_{min} is the minimum value in the signal matrix and L is the number of quantization level.

Then the input is divided into (L+1) levels with equal interval size ranging from m_{min} to m_{max}. When quantization step is done, then quantization value is fed to the next stage of compression.The three parameters mentioned above will be stored in the file so that quantization table can be created during the reconstruction process for dequantization.

3.2.2 Non Uniform Quantization

Different quantization steps will be applied in the speech dynamic range in non-uniform quantization. Uniform quantization is not a good strategy for distributions which significantly differ from uniform. If the distribution is non uniform, it is better to spend more quantization levels on more probable parts of the distribution and fewer quantization levels on less probable parts.

3.3 Encoding

The quantized data contains redundant information and it is a wastage of storage space. To overcome this problem, encoding schemes are exploited. Encoding can use any of the two compression techniques namely lossless or lossy.

Lossless compression algorithms rarely obtain a compression ratio larger than 3:1, while lossy compression allow compression ratio to range upto 12:1 and higher. For lossy algorithms, higher the compression but resulting lower audio quality. Examples of lossless compression are Huffman coding and run length coding.

3.3.1 Huffman Coding

Huffman encoding is a type of entropy encoding where the probabilities of occurrence of the symbols in the signals are computed. Then these symbols are arranged according to the probabilities of occurrence in descending order and build a binary tree and code word table.

The compressed speech signal can be reconstructed to form the original signal by decoding followed by dequantization and then performing the inverse transform methods. This would reproduce the original signal.

4. PERFORMANCE EVALUATION

A male spoken speech signal was taken as a test signal. Since the speech files were of short duration, the entire signal was decomposed at once without framing, The wavelets used in this paper are Haar and Daubechies (db2, db4, db6, db8, db10) wavelets [15]. Global threshold and uniform quantization were applied the transform coefficients. Huffman coding was employed to reduce the loss of quality.

To evaluate the overall performance of each transform techniques, several objective tests were made. Here performance measured in terms of CF, SNR, PSNR and NRMSE.

4.1 Compression Factor (CF)

Compression factor is also known as compression power used to quantify the reduction in data-representation size produced by a data compression algorithm. It is the ratio of the original signal to the compressed signal.

$$CF = \frac{\text{length of original signal}}{\text{length of compressed signal}}$$

4.2 Signal to Noise Ratio (SNR)

$$SNR = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right)$$

where σ_x^2 is the mean square of speech signal and σ_e^2 is the mean square difference between the original and reconstructed signal.

4.3 Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\frac{NX^2}{||x - x'||}$$

where N is the length of reconstructed signal, X is the maximum absolute square value of the speech signal x and ||x-x'|| is the energy difference between the original and reconstructed signal.

4.4 Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \sqrt{\frac{\sum_n[x(n) - x'(n)]^2}{\sum_n[x(n) - \mu_x(n)]^2}}$$

where x(n) is the speech signal, x'(n) is reconstructed speech signal and $\mu_x(n)$ is the mean of the speech signal. A summary of the performance of different transforms by using global threshold and uniform quantization is shown in table 1.

Table 1. Performance of Male Speech Signal

Technique	C.F	SNR (dB)	PSNR (dB)	NRMSE
FFT	2.1812	10.1821	30.0315	0.3268
DCT	3.0143	11.0137	30.4812	0.3127
Haar	4.2132	11.5629	32.6760	0.2905
Db2	4.1979	13.5463	34.9784	0.2229
Db4	4.2685	15.8458	35.7811	0.2032
Db6	4.2685	16.4659	35.3953	0.2011
Db8	4.2815	16.7117	35.6370	0.2066
Db10	4.3100	17.9800	35.9800	0.2016

It is shown that the highest compression is achieved by using DWT [16] decomposition as compare to FFT and DCT. Among different wavelets, Db10 wavelet provide the highest SNR, PSNR and lowest NRMSE, also Db10 [17] wavelet provides the highest greatest compression factor. The DCT shows its best result in terms of energy compaction, but its NRMSE is not acceptable.

The Haar wavelet transform is the simplest one to implement and it will be faster. It represents the same wavelet as Db1. However because of its discontinuity, it is not optimal to simulate a continuous signal. So Haar wavelet obtained the worst compression result as compared to other daubechies wavelet family.

By varying the value of truncation which is the percentage of coefficients that have been replaced by zero, the corresponding PSNR, SNR and NRMSE of the reconstructed signal is shown in fig 4, 5 and 6. Large percentage of coefficients are at low indices, therefore setting these values to zero would increase the compression factor. Setting the higher values to zero had little

effect on compression factor, because there are not many coefficients to affect.

It should be noted that truncation of coefficients as high as 90% still produces 30dB PSNR, 10Db SNR and 0.2 NRMSE. This means that only 10% of the transform coefficients are significant for that level of speech quality and discarding the rest of coefficients have little effect on speech quality. Hence signal of a certain quality can be reconstructed by choosing appropriate threshold.
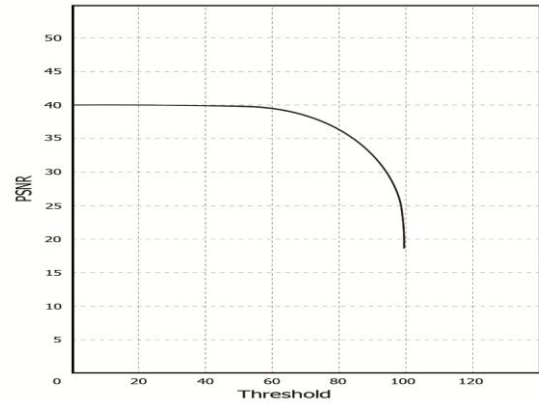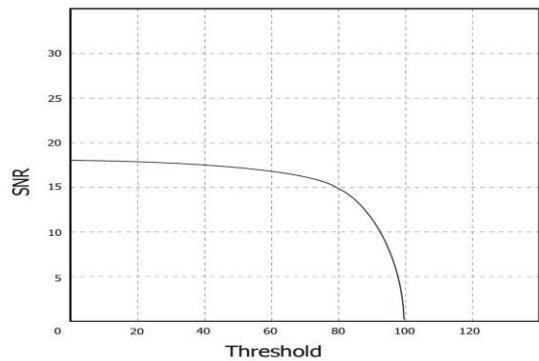

Fig 4: Threshold Vs PSNR

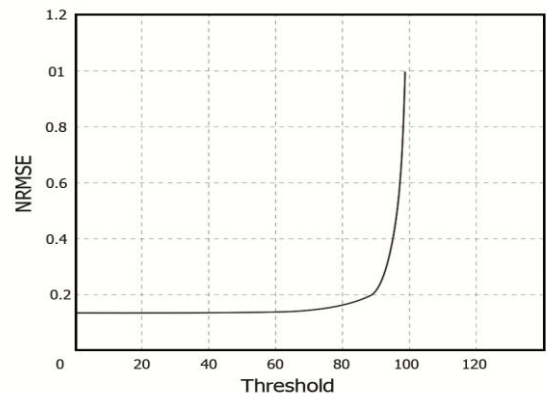**Fig 5: Threshold Vs SNR**

Fig 6: Threshold Vs NRMSE

5. CONCLUSION

Speech processing is an active area of research in the field of VOIP and mobile communication systems. The discrete wavelet transform performs very well in the analysis and processing of non-stationary speech signals. The greatest advantage of wavelet over other techniques is that the compression factor is

not constant and it can be varied while most other techniques have fixed compression factors. Discrete wavelet transform significantly improves the reconstruction of the compressed speech signal and also yields higher compression factor as compared to FFT and DCT. It is also observed that different wavelets have varying effects on the speech signal and also global threshold yields better results than the level dependent threshold method.

Much work must be done to improve the wavelet compression. More specifically, the scheme could improve by (i) finding the optimal mother wavelet and (ii) setting the truncation value which satisfy good compression factor and acceptable signal quality. These schemes can play very useful role in speech signal compression with reduced bitrates and good quality.

6. REFERENCES

[1] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.11, No.7, July 1989.

[2] R. S. H. Istepanian, A. Sungoor, and J. C. Nebel, "Linear Predictive Coding and Wavelet Decomposition for Robust Microarray Data Clustering", pp.4629-4632, Oct. 2007, IEEE.

[3] G. R. Redinbo, "Fault-Tolerant FFT Data Compression", Proceeding on Pacific Rim International Symposium on IEEE, Vol.29, pp. 2095-2105, Dec 2001

[4] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," IEEE Transactions on Computers, vol. C-32, pp. 90-93, Jan. 1974.

[5] F. W. Zaki, H. Hashish and S. H. behir, "Speech Compression Using wavelet Transform", Proceedings of the Eighteenth National Radio Science Conference, 2001, Vol. 2, pp 467-474, March 27-29, 2001.

[6] W. Chong and J. Kim, "Speech and Image Compression by DCT, Wavelet, and Wavelet Packet", Proceeding in Information, Communication and Signal processing IEEE, Vol.3, pp.1353-1357,Aug 2002.

[7] P.Noll, "Wideband Speech and Audio Coding," IEEE Communications Magazine, pp. 34-44, Nov. 1993.

[8] Q. Memon, T. Kasparis, "Transform Coding of Signals Using Approximate Trigonometric Expansions". Journal of Electronic Imaging, Vol. 6, No. 4, October 1997, pp. 494-503.

[9] N. M. Hosny S. H. El-Ramly and M. H. El-Said, "Noval Techniques For Speech Compression Using Wavelet Transform", Eleventh Int. Conf. on Microelectronics (ICM99), pp.225-229, 2002.

[10] Harmanpree Kaur and Ramanpreet Kaur, "Speech compression and decompression using DCT and DWT", International Journal Computer Technology &Applications,Vol 3 (4), 1501-1503 IJCTA | July-August 2012.

[11] Abdul Mawla M. A. najih, A. R. bin Ramli, V. Prakash, and Syed A. R., "Speech Compression Using Discreet Wavelet Transform", Proceedings of 4

[12] K. Ramchandran, M. Vetterli, and C. Herley, "Wavelet, Subband Coding, and Best Bases", Proceeding of the IEEE, Vol. 84, No. 4, April 1996.

[13] G. Rajesh1, A. Kumar and K. Ranjeet, "Speech Compression using Different Transform Techniques", International Conference on Computer & Communication Technology (ICCCT)-2011

[14] N. Junejo, N. Ahmed, M. Ali Unar, and A. Q. K Rajeput, "Speech and Image Compression Using Discrete Wavelet Transform", Advances in Wired and Wireless Communication, IEEE/Sarnoff Symposium, pp.45-48, 09 May 2005.

[15] 0.Rioul and M.Vetterli, "Wavelets and Signal Processing", IEEE Signal Process. Mag. Vo1.8, pp.14-38, Oct. 1991.

[16] E.B. Fgee, W.J. Phillips, W. Robertson, "Comparing Audio Compression using Wavelets with other Audio Compression Schemes", IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Edmonton, Canada, pp. 698-701, 1999.

[17] LDaubechies, "The Wavelet Transform, Time Localization and Signal Analysis, IEEE Trans. Inform. Theory, Vol. IT-36, pp.951-1006, Sept. 1980.

[18] "A Tutorial of the Wavelet Transform" by Chun-Lin, Liu in February 23,2010.

[19] L.Rabiner, B. Juang , hndamental of Speech Recognition, Prentice Hall 1993.

[20] Gilbert Strang, Truong Nguyen. Wavelets and Filter Banks, Wellesley Cambridge Press 1996.