

Audio and Speech Compression Using DCT and DWT Techniques

M. V. Patil¹, Apoorva Gupta², Ankita Varma³, Shikhar Salil⁴

Asst. Professor, Dept.of Elex, Bharati Vidyapeeth Univ.Coll.of Engg, Pune, Maharashtra, India¹

B.Tech. student, Dept.of Elex, Bharati Vidyapeeth Univ.Coll.of Engg, Pune, Maharashtra, India^{2,3,4}

Abstract: Audio compression has become one of the basic technologies of the multimedia age. The change in the telecommunication infrastructure, in recent years, from circuit switched to packet switched systems has also reflected on the way that speech and audio signals are carried in present systems. In many applications, such as the design of multimedia workstations and high quality audio transmission and storage, the goal is to achieve transparent coding of audio and speech signals at the lowest possible data rates. In other words, bandwidth cost money, therefore, the transmission and storage of information becomes costly. However, if we can use less data, both transmission and storage become cheaper. Further reduction in bit rate is an attractive proposition in applications like remote broadcast lines, studio links, satellite transmission of high quality audio and voice over internet.

Keywords: DCT (Discrete cosine transform), DWT (Discrete wavelet transform), Quantization, Compression Factor (CF), Signal to Noise ratio (SNR), Peak Signal to Noise Ratio (PSNR)

I. INTRODUCTION

Speech is very basic way for humans to convey information. The main objective of Speech is communication. Speech can be defined as the response of vocal track to one or more excitation signal. Huge amount of data transmission is very difficult both in terms of transmission and storage. Speech Compression is a method to convert human speech into an encoded form in such a way that it can later be decoded to get back the original signal. Compression is basically to remove redundancy between neighboring samples and between adjacent cycles. Major objective of speech compression is to represent signal with lesser number of bits. The reduction of data should be done in such a way that there is acceptable loss of quality.

TYPES OF COMPRESSION

There are mainly two types of compression techniques - Lossless Compression and Lossy Compression

1.1. Lossless Compression

It is a class of data compression algorithm that allows the exact original data to be reconstructed from the exact original data to be reconstructed from the compressed data. It is mainly used in cases where it is important that the original signal and the decompressed signal are almost same or identical. Examples of lossless compression are Huffmann coding.

1.2 Lossy Compression

It is a data encoding method that compresses data by removing some of them. The aim of this technique is to minimize the amount of data that has to be transmitted. They are mostly used for multimedia data compression. The rest of the paper is organized as follow, section 2 gives the Theoretical background about the speech compression schemes. The speech compression techniques are described in section 3& Section 4 evaluates the performance of the proposed technique followed by the conclusion.

II. LITERATURE BACKGROUND

HarmanpreetKaur and RamanpreetKaur proposed speech compression method using different transform techniques. The signal is compressed by DWT technique afterward this compressed signal is again compressed by DCT and then this compressed signal is decompressed using DWT technique. Performance of speech signal is measured on the basis of Peak Signal to Noise Ratio and Normalized Root Mean Square Error (NRMSE) by using different filters of wavelet family. [1]

Jalal Karam and RautSaad discussed the effects of different compression constraint & schemes to achieve high compression ratio and acceptable signal to noise ratio. Flexibility is obtained by observing Wavelet Used, decomposition Level, Compression Ratio, Frame Size, Measured Parameters & types Of Threshold used. High compression ratios were achieved with acceptable SNR. The resultant signal is compared with Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE). [2]

SmitaVatsa and Dr.O.P.Sahu implemented various speech compression techniques. Transform coding is based on compression of signal by removing redundancies present in it. It is a process of transforming signal into compressed or compact form so that the signal could be stored with lesser bandwidth. In this paper DCT and DWT based speech compression techniques are implemented with Run Length Encoding, Huffmann Encoding. Reconstructed signals are compared using factors like Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE). [3]

HatemElaydi and Mustafi I. Jaber and Mohammed B. Tanboura gave a new lossy algorithm to compress speech signal using DWT techniques. Due to growth of multimedia technology over the past decade, demand for digital information has increased. The only way to overcome this situation is to compress the information signal by removing the redundancies present in them. The compression ratio can be easily varied by using wavelets while other methods have fixed compression ratios. [4]

OathmanO.Khalifa, SeringHabib Harding and Aisha-Hassan described the importance and need for audio compression. Audio compression has become one of the most basic technologies of this period. [5]

2.1 Techniques for speech compression: Speech compression is classified into three categories:-

- **Waveform Coding**

The signal that is transmitted as input is tried to be reproduced at the output which would be very similar to the original signal.

- **Parametric coding**

In this type of coding the signals are represented in the form of small parameters which describes the signals very accurately. In parametric extraction method a preprocessor is used to extract some features that can be later used to extract the original signal.

- **Transform Coding**

This is the coding technique that we have used for our paper. In this method the signal is transformed into frequency domain and then only dominant feature of signal is maintained. In transform method we have used discrete wavelet transform technique and discrete cosine transform technique. When we use wavelet transform technique, the original signal can be represented in terms of wavelet expansion. Similarly in case of DCT transform speech can be represented in terms of DCT coefficients. Transform techniques do not compress the signal, they provide information about the signal and using various encoding techniques compressions of signal is done. Speech compression is done by neglecting small and lesser important coefficients and data and discarding them and then using quantization and encoding techniques.

Speech compression is performed in the following steps.

- *Transform technique*
- *Thresholding of transformed coefficients*
- *Quantization*
- *Encoding*

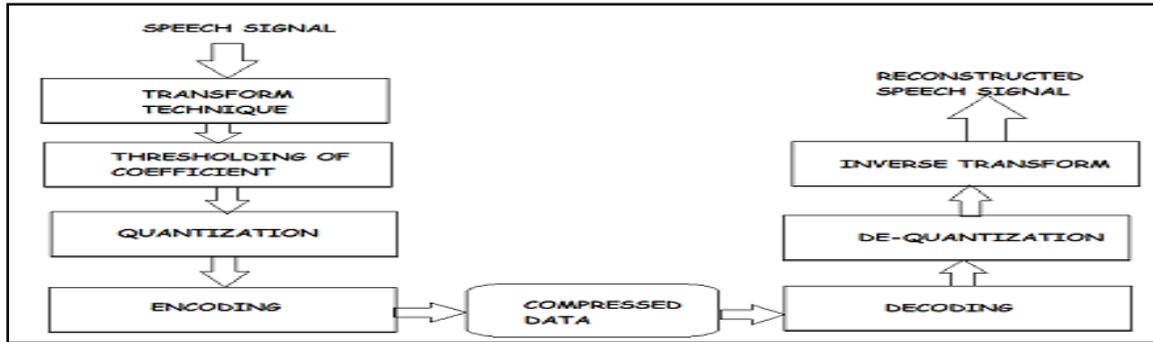


Fig-1:Block diagram of speech compression

A.) Transform Technique

DCT and DWT methods are used on speech signal. Using DCT, reconstruction of signal can be done very accurately; this property of DCT is used for data compression. Localization feature of wavelet along with time frequency resolution property makes DWT very suitable for speech compression. The main idea behind signal compression using wavelets is linked primarily to the relative scarceness of the wavelet domain representation of signal.

1. Discrete Wavelet Transform

A discrete wavelet transform can be defined as a ‘small wave’ that has its energy concentrated in time, and it provides a tool for the analysis of transient, non-stationary or time varying phenomenon. It has oscillating wave like property. Wavelet is a waveform of limited duration having an average value zero. They are localized in space. Wavelet transform provides a time-frequency representation of the signal. In DWT, the signal is decomposed into set of basic functions also known as ‘WAVELETS’. Wavelets are obtained from a single MOTHER WAVELET by delay and shifting.

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right)$$

Where ‘a’ is the scaling parameter and ‘b’ is the shifting parameter. DWT uses multiresolution technique to analyze different frequencies. In DWT, the prominent information in the signal appears in the lower amplitudes. Thus compression can be achieved by discarding the low amplitude signals. Fig.2 shows the compression system design.

2. Discrete Cosine Transform

Discrete Cosine Transform can be used for speech compression because of high correlation in adjacent coefficient. We can reconstruct a sequence very accurately from very few DCT coefficients. This property of DCT helps in effective reduction of data.

DCT of 1-D sequence x (n) of length N is given by

$$X(m) = \left[\frac{2}{N} \right]^{1/2} C_m \sum_{n=0}^{N-1} x(n) \cos \left[\frac{(2n+1)m\pi}{2N} \right]$$

Where m=0, 1, -----, N-1

The inverse discrete cosine transform is

$$X(n) = \left[\frac{2}{N} \right]^{1/2} \sum_{m=0}^{N-1} C_m x(m) \cos \left[\frac{(2n+1)m\pi}{2N} \right]$$

In both equations Cm can be defined as

$$C_m = (1/2)^{1/2} \text{ for } m=0$$

1 for $m \neq 0$

It expresses a finite sequence of data points in terms of sum of cosine function oscillating at different frequencies. They are very common encoding technique for audio track compressions. It is very similar to DFT, but the only difference is that the output vector is approximately twice as long as the DFT output. They are used in JPEG image compressions, MJPEG and many other video compressions.

B.) Thresholding

After the coefficients are received from different transforms, thresholding is done. Very few DCT coefficients represent 99% of signal energy; hence Thresholding is calculated and applied to the coefficients. Coefficients having values less than threshold values are removed.

C.) Quantization

It is a process of mapping a set of continuous valued data to a set of discrete valued data. The aim of quantization is to reduce the information found in threshold coefficients. This process makes sure that it produces minimum errors. We basically perform uniform quantization process.

D.) Encoding

We use different encoding techniques like Run Length Encoding and Huffmann Encoding. Encoding method is used to remove data that are repetitively occurring. In encoding we can also reduce the number of coefficients by removing the redundant data. Encoding can use any of the two compression techniques, lossless or lossy. This helps in reducing the bandwidth of the signal hence compression can be achieved. The compressed speech signal can be reconstructed to form the original signal by DECODING followed by DE-QUANTIZATION and then performing the INVERSE-TRANSFORM methods. This would reproduce the original signal.

2.2 WAVELET BASED COMPRESSION TECHNIQUES

Wavelets concentrate speech signals into a few neighbouring coefficients. By taking the wavelet transform of a signal, many of its' coefficients will either be zero or have negligible magnitudes. Data compression can then be done by treating the small valued coefficients as insignificant data and discarding them. Compressing a speech signal using wavelets involves the following stages.

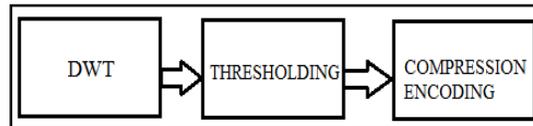


Fig.2 :: Compression System Design

A.) Choice of wavelets

Choosing mother-wavelet function which is used in designing high quality speech coders is of prime importance. Choosing a wavelet having a compact support in time and frequency in addition to a significant number of vanishing moments is important for wavelet speech compressor. Different criteria can be used in selecting an optimal wavelet function. The objective is to minimize the error variance and maximize signal to noise ratio. They can be selected based on the energy conservation properties. Better reconstruction quality is provided by wavelets with more vanishing moments, as they introduce lesser distortion and concentrate more signal energy in neighbouring coefficients. However the computational complexity of DWT increases with the number of vanishing moments. Hence it is not practical to use wavelets with higher number of vanishing moments. Number of vanishing moments of a wavelet indicates the smoothness of a wavelet function and also the flatness of the frequency response of the wavelet filters. Higher the number of vanishing moments, faster is the decay rate of wavelet coefficients. It leads to a more compact signal representation and hence useful in coding applications. However, length of the filters increases with the number of vanishing moments and the hence complexity of computing the DWT coefficients increases.

B.) Decomposition of wavelets

Wavelets decompose a signal into different resolutions or frequency bands. Signal compression is based on the concept that selecting small number of approximation coefficients and some of the detail coefficients can represent the signal components accurately. Choosing a decomposition level for the DWT depends on the type of signal being used or parameters like entropy.

C.) Truncation of coefficients

Compression involves truncating wavelet coefficients below threshold. Most of the speech energy is high-valued coefficient. Thus the small valued coefficients can be truncated or zeroed and can then be used for reconstruction of the signal. This compression technique provided lesser signal-to-noise ratio. Two different methods are available for the calculation of thresholds.

- **Global Thresholding-** It takes the wavelet expansion of the signal and keeps the largest absolute value coefficient. In this we manually set a global threshold. Hence only a single parameter needs to be selected in this case.
- **Level Thresholding-** It applies visually determined level dependent thresholds to each of the decomposition level in the wavelet transform.

D.) Encoding coefficients

Signal compression is achieved by first truncating small-valued coefficients and then encoding these coefficients. High-magnitude coefficients can be represented by storing the coefficients along with their respective positions in the wavelet transform vector. Another method for compression is to encode consecutive zero valued coefficient with two bytes. One byte indicates the sequence of zeros in the wavelet transforms vector and the second byte represents the number of consecutive zeros. For further data compression a suitable bit-encoding format can be used. Low bit rate representation of signal can be achieved by using an entropy coder like Huffman coding.

E.) Calculating threshold

Two different thresholding techniques are used for the truncation of coefficients i.e. global thresholding and level thresholding.

The design of flow model is depicted in Figure-3

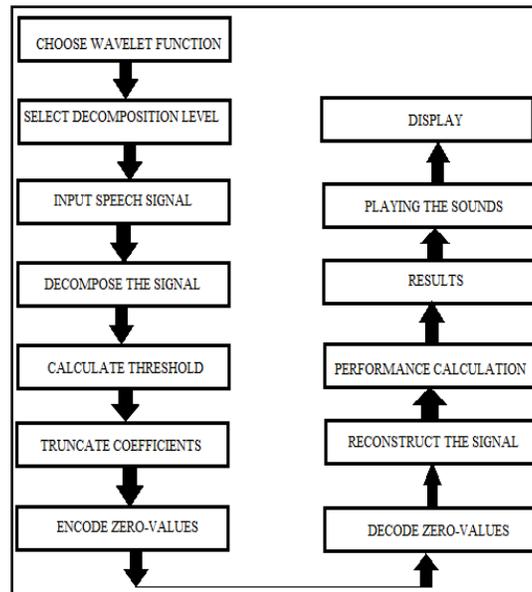


Fig-3 : Design Flow of Wavelet Based Speech Coder

Global Thresholding process retains the largest absolute value coefficients. Global thresholds can be calculated by setting the percentage of coefficients to be truncated.

Level dependent thresholds are calculated using the Birge-Massart strategy. This thresholding process is based on the approximation result from Birge and Massart.

F.) Encoding zero value functions

In this method, consecutive zero valued coefficients are encoded with two bytes. One byte specifies the starting string of zeros and the second byte keeps record of the number of successive zeros. This encoding method provides a higher compression ratio.

2.3 DCT BASED COMPRESSION TECHNIQUE

The given sound file is read. The vector is divided into smaller frames and arranged into matrix form. DCT operation is performed on the matrix. DCT operation is performed and the elements are sorted in their matrix form to find components and their indices.

The elements are arranged in descending order. After the arrangement has been done, a Threshold value is decided. The coefficients below the threshold values are discarded.

Hence reducing the size of the signal which results in compression. The data is then converted back into the original form by using reconstruction process. For this we perform IDCT operation on the signal. Now convert the signal back to its vector form. Thus the signal is reconstructed.

The fig.4 shows the DCT compression and decompression process.

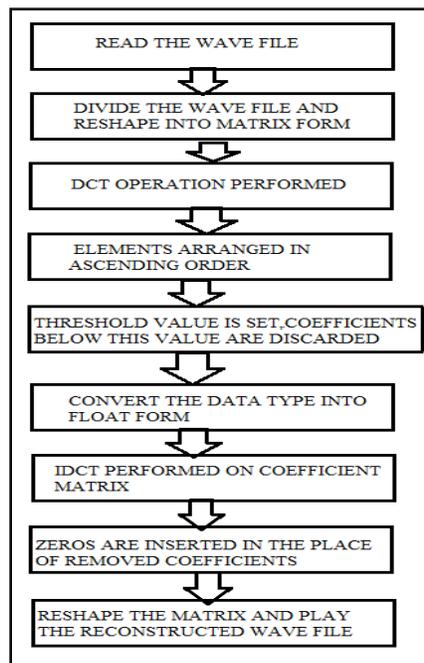


Fig.4: DCT based speech coder design flow

III. PERFORMANCE EVALUATION

To evaluate the overall performance of the proposed audio compression scheme, several objective tests were made. To measure the performance of the reconstructed signal, various factors such as compression factor, Signal to noise ratio, PSNR & mean square error are taken into consideration.

3.1. Compression Factor (CF)

We take into account all the values that would be required to completely represent the signal.

$$\text{COMPRESSION FACTOR} = \frac{\text{LENGTH OF ORIGINAL SIGNAL}}{\text{LENGTH OF COMPRESSED SIGNAL}}$$

3.2. Signal to Noise Ratio (SNR)

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

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

3.3. Peak Signal to Noise Ratio (PSNR)

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

Where N is the length of reconstructed signal, X is the maximum absolute square value of signal x and $\|x-x'\|^2$ is the energy of the difference between the original and reconstructed signal.

3.4. Normalized Root Mean Square Error (NRMSE)

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

Here, X(n) is the speech signal, x'(n) is reconstructed speech signal and $\mu_x(n)$ is the mean of speech signal.

3.5. Retained Signal Energy:

$$RSE = \frac{\|x'(n)\|^2}{\|x(n)\|^2} \times 100$$

Here, $\|x'(n)\|^2$ and $\|x(n)\|^2$ represent energy of reconstructed and original speech signal.

RSE is the amount of energy retained in the compressed signal as a percentage of the energy of original signal.

The results for Compression factor, Signal to Noise ratio, PSNR & Mean square error for the "HEY, HOW ARE YOU" (1) and "APPLE" (2) signal using the DCT based compression are summarized in table 1.

Signal	CF	SNR(dB)	PSNR(dB)	MSE
Audio 1	0.2639	31.83	45.21	0.02990
Audio 2	0.3433	32.46	47.17	0.05327

Table 1: Results of DCT based technique in terms of CF, SNR, PSNR& MSE

The results for DWT based compression and its corresponding wavelets for the same audio samples are listed in Table 2.

Signal	Technique	CF	SNR(DB)	PSNR(DB)	MSE
Audio 1	DB2	0.0512	19.76	35.49	0.02
Audio 2	DB2	0.0783	43.42	44.15	0.06
Audio 1	DB10	0.0461	20.04	35.86	0.08
Audio 2	DB10	0.0636	45.56	45.57	0.07
Audio 1	HARR	0.0711	17.33	34.17	0.19
Audio 2	HARR	0.0945	42.67	42.83	0.20

Table 2: Results of DWT based tech in terms of CF,SNR,PSNR,& MSE

IV. CONCLUSION

A simple discrete wavelet transform & DCT based audio compression scheme presented in this paper. It is implemented using MATLAB . Experimental results show that in general there is improved in compression factor & signal to noise ratio with DWT based technique. It is also observed that Specific wavelets have varying effects on the speech signal being represented.

REFERENCES

- [1] HarmanpreetKaur and RamanpreetKaur, "Speech compression and decompression using DCT and DWT", International Journal Computer Technology & Applications, Vol 3 (4), 1501-1503 IJCTA | July-August 2012
- [2] Jalal Karam and RautSaad, "The Effect of Different Compression Schemes on Speech Signals", International Journal of Biological and Life Sciences, 1:4, 2005
- [3] O. Rioul and M. Vetterli, "Wavelets and Signal Processing", IEEE Signal Process. Mag. Vol 8, pp. 14-38, Oct. 1991..
- [4] HatemElaydi and Mustafi I. Jaber and Mohammed B. Tanboursa, "Speech compression using Wavelets", International Journal for Applied Sciences, Vol 2, 1-4, Sep 2011
- [5] Othman O. Khalifa, SeringHabib Harding & Aisha-Hassan A. Hashim "Compression using Wavelet Transform" in Signal Processing: An International Journal, Volume (2) : Issue (5).