

Analysis and Comparison of Audio Compression Using Discrete Wavelet Transform

Vimal Kishore Yadav¹, Alok Jain², Lenka Bhargav³

Assistant Professor, ASET, Amity University, Madhya Pradesh, India^{1,2}

Student, ASET, Amity University, Madhya Pradesh, India³

Abstract: Speech compression is the technique of converting human speech into an efficiently encoded format that can later be decoded to produce a close approximation of the original signal. The merits of the compression technique are reduction in storage space, bandwidth, transmission power and energy. An efficient algorithm Discrete Wavelet Transform (DWT) is employed for decomposition of original signal into wavelets coefficients at different scales and positions and these coefficients are truncated to perform encoding and decoding. The compression technique used in this paper is better than other earlier coding techniques like μ -law coding, code excited linear predictive coding. Speech compression plays a prominent role in speech signal processing such as satellite communications, internet communications, transmission of biomedical signals and other applications.

Keywords: Wavelet Transform, speech compression, Filtering, Truncation.

I. INTRODUCTION

Speech is a weapon to convey the information from one person to other which is limited to a bandwidth of 4 kHz. The growth of multimedia communication demands for efficient use of transmission bandwidth, storage space and power. To overcome the consequences we need to compress data and this paper has a novice look in analyzing and compressing the speech signal. PCM was the primary audio technique employed in the earlier times for encoding but it doesn't provide any mechanism for redundancy removal. Over the last few researchers have propounded several efficient algorithms based on transformation such as Fourier Transform (FT), Wavelet Transform (WT), and Discrete Cosine Transform (DCT). Fast Fourier transform (FFT) has been popular compression algorithm in the earlier days but it is based on the assumption that signal is stationary. After invention of Short Time Fourier Transform (STFT), it resolved some of the issues but not completely. Most of signals present in the world are non-stationary and for better results this should be analyzed without any assumption where we take an assumption that signals are linear and stationary for Fourier analysis.

The use of wavelets has been revolutionized in signal analysis which is used to analyze the linear and non-stationary signal. Speech is a non-stationary random process due to the time varying nature and abrupt change in frequencies of the human speech production system. Non-stationary signals are the signals characterized by transient changes in their frequencies. The important localization feature of wavelets, along with their resolution properties in time-frequency makes them well suitable for coding speech signals. Speech compression which is very necessary for the present world due to advancement in multimedia technology since the bandwidth is limited.

The outline of the paper is as follows: section II covers the introduction of wavelet, wavelet families; DWT. Section III covers speech compression using DWT, decomposition

process and encoding. Section IV covers performance measures; Section V covers results, Section VI covers conclusion and section VIII covers references.

II. INTRODUCING WAVELET

Wavelet is one of the recent developments to overcome the limitations of Fourier transform of signal analysis which has the special ability to examine signal simultaneously in both time and frequency. Wavelet analysis is breaking up of an original signal into scaled and shifted versions called the mother wavelet. Wavelet Transform has emerged as a recent powerful mathematical tool in many areas of science and technology, more so in the field of signal processing.

1. PROPERTIES OF MOTHER WAVELET:

The important and necessary properties of wavelets are the admissibility and the regularity conditions and these are the properties which gave wavelets their name [1].

Consider a real valued continuous-time function $\psi(t)$

Property 1: Admissibility condition 1,

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$

Property 2: Admissibility condition 2,

$$\int_{-\infty}^{\infty} \psi(t) dt = 0$$

Property 3: A zero at the zero frequency also means that the average value of the wavelet in the time domain must be zero,

$$|\psi|_{\omega=0}^2 = 0$$

There are infinity of functions which satisfy the above properties, amongst which 'Haar wavelet' is the simplest of them.

2. WAVELET FAMILIES:

Wavelet families are the important tools for analyzing a signal. Every wavelet has its own properties and choosing a wavelet for analyzing a signal depends upon the type of application we perform. Within the family of wavelets, wavelet often classified by the number of vanishing moments N which is written as $D(N)$. We can see Daubechies $D10$ with vanishing points 10 from figure (II.2.a)

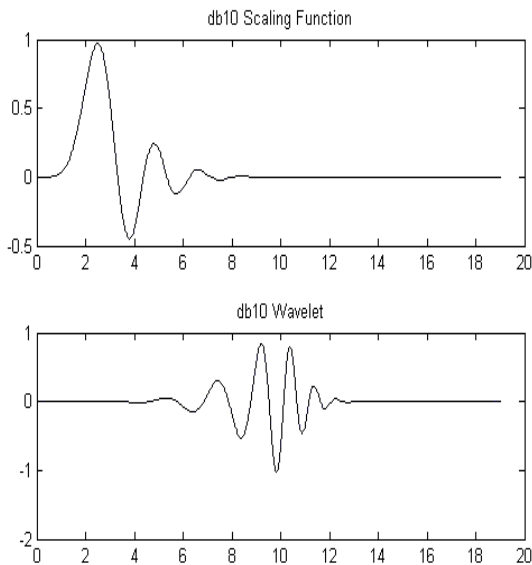


Fig (II.2.a) Daubechies (D10) scaling and wavelet

3. DISCRETE WAVELET TRANSFORM:

DWT is an important tool for analysis of audio signal, which is non-stationary. DWT consists of coefficients of expansion of original signal with respect to a basis $\psi_{j,k}(t)$ of which is a dilated version of function ψ called mother wavelet [2].

Expression for DWT is as follows:

$$\Psi_{j,k}(t) = \frac{1}{s_0^j} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right), j, k \in Z$$

It is the decomposition of signal into approximation and detailed coefficients using cascaded filter banks consisting of low and high-pass filters. The output of the low-pass filter provides approximation coefficients and high-pass filter gives detailed coefficients.

III. SPEECH COMPRESSION USING DISCRETE WAVE TRANSFORMS (DWT):

1. CHOICE OF WAVELETS

The primary importance in designing speech coders is to choose the mother wavelet. Wavelet has to be chosen such that it has a compatibility both in time and frequency domain as well as which has the significant number of vanishing moments for optimum audio coding. Wavelets with large number of vanishing are useful for this audio compression method because wavelet with large number of vanishing moments can give a precise specification of pass bands of each sub-band in wavelet decomposition and compactness in signal representation. Several different

criteria can also be used in selecting an optimal wavelet function for best results. The main objective of the paper is to minimize reconstructed error variance and maximize signal to noise ratio (SNR).

In [3] it is shown Daubechies $D20$, $D12$, $D10$ or $D8$ wavelets, all concentrate more than 96% of the signal energy in the Level 1 approximation coefficients. The reconstruction of the original signal increases with increase in vanishing moments of the analysis. However the computational complexity increases with the increase of vanishing moments and hence it is advised for signal coding applications that it is not practical to use wavelets with an arbitrarily high number of vanishing moments [4].

2. WAVELET DECOMPOSITION:

From the figure (Fig III.2.a) we can define wavelet decomposition as the decomposition of original signal S into different frequency bands, and this decomposition is carried out by choosing an appropriate wavelet function with required vanishing moments and level of decomposition [5]. Wavelets works by decomposing a signal based on the concept that selecting suitable low and high pass filter banks for passing low frequencies and high energy coefficients called approximation coefficients (A) and high frequency and low energy called detailed coefficients (D). The original signal passes through two complementary filters and emerges as two signals with coefficients A and D. The decomposition level depends on the type of signal being used or some other suitable criterion such as entropy. The decomposition up to level 5 is generally suitable for speech signal processing.

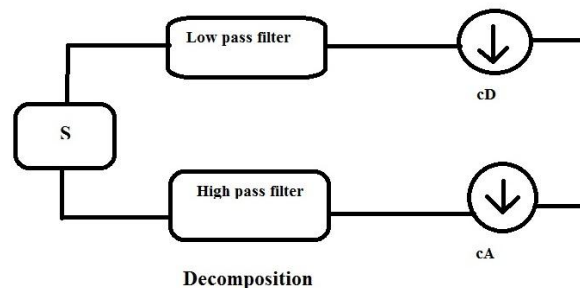


Fig. (III.2.a)

3. FILTERING: APPROXIMATION AND DETAILS

Frequency is one of the important parameters to Characterize a signal. For many signals, the low-frequency content is the most important part which signifies the identity of the signal. The high-frequency content, imparts details of the signal. Consider a speech signal, remove the high-frequency components, the voice sounds different, but one can guess what is being said. However, if you remove enough of the low-frequency components, you hear gibberish. The original signal, S , passes through two matching filters and emerges as two signals say A and D. For suppose that the original signal S consists of 2400 samples of data then the resulting signal, consists each of

2400samples, a total of 4800 samples which results in increase in samples than the original. To overcome the increase in sampling we perform down sampling to produce two sequences called cA and cD of coefficients 1200 each. The detailed coefficients are small and consist mainly of a high-frequency noise, on the other hand while the approximation coefficients cA contains much less noise than the original signal does. For reconstruction of original signal S from approximation and details coefficients we use up sampling and filtering method. The reconstruction involves up sampling to increase the number of samples as that of the original signal S. Up-sampling is the process of lengthening a signal component by inserting zeros between samples.

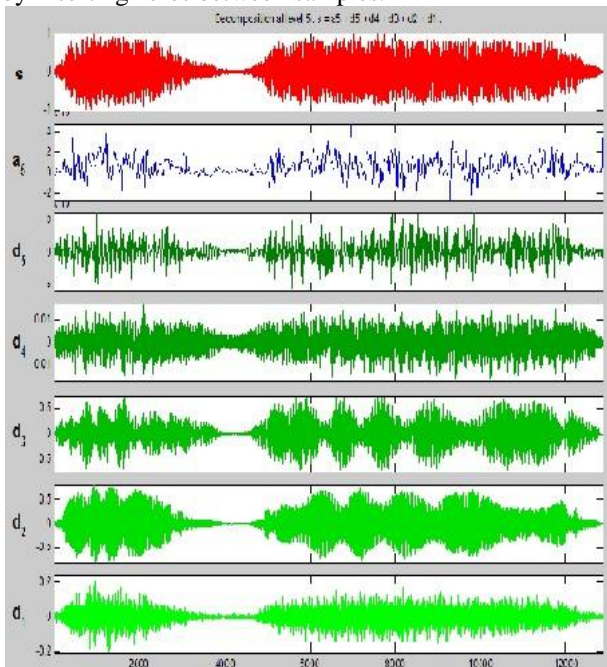


Fig. (III.3.a) Decomposition of demo signal (train.mat) using d10 at level 5

4. TRUNCATION OF COEFFICIENTS:

After calculating the wavelet transform of the speech signal S, compression involves truncating wavelet coefficients below a threshold. From the experiments conducted, it has been found that most of the coefficients have small magnitudes through which we can infer more than 85% of the wavelet coefficients were found to be insignificant, and their truncation to zero made an unnoticeable difference to the signal. This indicates that most of the speech energy is concentrated in the high-valued coefficients, which are limited. Thus the small valued coefficients can be truncated or zeroed and then be used to reconstruct the original signal. There are two different approaches available for calculating the thresholds.

a) Global Threshold:

This thresholding method involves keeping the high value coefficients of the signal which are a few in numbers and can be done manually by setting the global threshold.

b) Level Dependent Threshold:

In this method we apply visually determined level dependent thresholds to each decomposition level of signal.

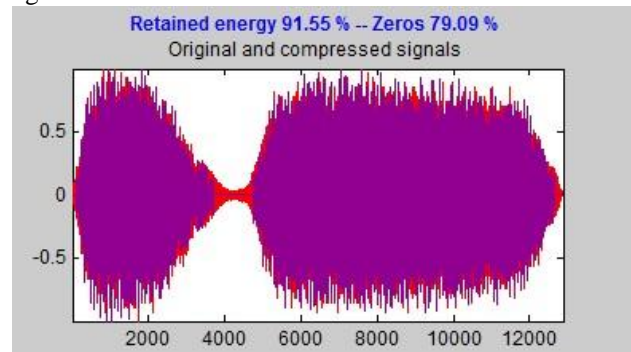


Fig. (III.4.a) Retained Energy 91.55% - Zeros 79.09%

4. ENCODING ZERO VALUED COEFFICIENTS:

Before encoding the signal efficiently signal compression is accomplished by first truncating small-valued coefficients.

[6] Our approach to compression is to encode consecutive zero valued coefficients with two bytes. One byte to indicate a sequence of zeros in the wavelet transform vector and the second byte represent the number of consecutive zeros.

IV. PERFORMANCE MEASURES

These are some parameters which evaluate the quality of signal after compression and reconstruction:

a. Compression ratio: Compression factor is defined as the ratio of the original signal to the compressed signal.

$$CR = \frac{\text{Length}(h(x(n)))}{\text{Length}(h(r(n)))}$$

b. Retained signal energy: This indicates the amount of energy retained in the compressed signal as a percentage of the energy of original signal.

$$RSE (\%) = \frac{\|x(n)\|^2}{\|r(n)\|^2} * 100$$

c. Signal to noise ratio (SNR): This value gives the quality of reconstructed signal.

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

d. Percentage of zero coefficients:

100*(% of current decomposition)/% of coefficient

V. APPLICATION OF WAVELETS

Signal de-noising, Signal compression, ECG Heart Beat rate monitoring, De-noising of satellite images, image processing, mathematical applications.

VI. RESULT

The results of the Discrete Wavelet Transform in terms of compression factor and signal to noise ratio are comparable with the other techniques such as code excited linear predictive coding (CELP), μ -law compression, RPE

codec's. The use of DWT makes the mathematical computations less complex and increase in speed in calculations over CWT and earlier methods.

VII. CONCLUSION

The compression of speech signal has become an important factor as there is an increase in multimedia communication. The Discrete Wavelet Transform performs very well in the compression of recorded speech signals. For real time speech processing, however, its performance is not as good as that of recorded signal. Some other criteria should be selected for real time signals like with less vanishing moments.

REFERENCES

- [1]. W.Kinsner and A.Langi "*speech and image signal compression with wavelets*".W.Kinsner and A.Langi.
- [2] Amandeep Kour, Vimal kishore Yadav, Vikas maheshwari, Deepak Prashar, "A Review on Image Processing" International Journal of Electronics Communication and Computer Engineering, vol. 4, Issue 1 ISSN(online):2249-071X, ISSN(print):2278- 4209, 2013.
- [3] Fgee, E.B., Phillips, W.J. and Robertson, W. "*Comparing Audio Compression using Wavelets with other Audio Compression Schemes*", IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Edmonton, Canada, 1999, pp. 698-701.
- [4] Amandeep Kour, Vimal kishore Yadav, Vikas maheshwari, Deepak Prashar, "*Web Mining in Soft Computing Relevance and Future Directions Delay*", International Journal of Electronics Communication and Computer Engineering, vol. 4, Issue 1 ISSN(online):2249-071X, ISSN(print):2278-4209, 2013.
- [5] Vimal Kishore Yadav, Pratiksha Singh Gaur, Sucheta Yadav, "*Impending Form Interpretations for Delay to Ramp and Step Input On-Chip VLSI RLC Annex* ", "International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE)", vol. 3, Issue 11, ISSN (Online) : 2278-1021,ISSN (Print) : 2319-5940,2014
- [6] Amandeep Kour, Vimal kishore Yadav, Vikas maheshwari, Deepak Prashar, "*Face Recognition Using Template Matching*", " IEEE International Conference on Signal and Image Processing