Speech Compression Using Discrete Wavelet Transform and Discrete Cosine Transform

Smita Vatsa,

M. Tech (ECE Student)
Department of Electronicsand
Communication Engineering
NIT Kurukshetra, India

Dr. O. P. Sahu

Professor
Department of Electronics and
Communication Engineering
NIT Kurukshetra, India

Abstract

Aim of this paper is to explain and implement transform based speech compression techniques. Transform coding is based on compressing signal by removing redundancies present in it. Speech compression (coding) is a technique to transform speech signal into compact format such that speech signal can be transmitted and stored with reduced bandwidth and storage space respectively .Objective of speech compression is to enhance transmission and storage capacity. In this paper Discrete wavelet transform and Discrete cosine transform based speech compression techniques are implemented with Run length encoding, Huffman encoding and Run length encoding followed by Huffman encoding. Reconstructed speech obtained from implementation are compared on the basis of Compression Factor(CF), Signal to noise ratio (SNR), Peak signal to noise ratio (PSNR), Normalized root mean square error (NRMSE), Retained signal energy (RSE). Results shows that Discrete wavelet transform with RUN length encoding followed by Huffman encoding gives higher compression with intelligible reconstructed speech.

1.Introduction

Objective of speech is communication whether face to face or cell phone to cell phone. A huge amount of data is a big issue for transmission or storage. Speech compression is the technology of converting human speech into an efficiently encoded representation that can later be decoded to produce a close approximation of the original signal. Major objective of speech compression is to represent speech with less or few number of bits with level of quality. Speech coding means representing speech signal in an efficient way for transmission and storage such a way it can be reproduced with desired level of quality. Main approaches of speech compression used today are waveform coding, transform coding and parametric coding. Waveform coding attempts to reproduce input signal waveform at the output. In transform coding at the beginning of procedure signal is transformed into frequency domain, afterwards only dominant spectral features of signal are maintained. In parametric coding signals are represented through a small set of parameters that can describe it accurately. Parametric coders attempts to produce a signal that sounds like original speechwhether or not time waveform resembles the original.

By removing redundancy between neighbouring samples signal can be compressed. In this paper we have implemented compression technique in two step, in 1st step a transform function is applied on speech signal to get result with a new set of data with smaller values and more repetition, 2nd step is coding(compression) step, this step willrepresent the data set in its minimal form by using encoding techniques such as Run Length encoding, Huffman encoding, Run lengthencoding followed by Huffman encoding. Performance measurescompression factor(CF),signal to noise ratio(SNR),peak signal to noise ratio(PSNR),normalized root mean square error(NRMSE), retained signal energy (RSE) is measured for reconstructed speech obtained from DWT and DCT based speech compression techniques. A comparative analysis of performance of transform technique based speech compression systems for 3 different encoding methods is done. In this paper a brief introduction of speech compression technique is presented in section 2, section 2 discusses Transform techniques, in section 3 implementation of speech compression based different transform technique with several encoding methods is explained, section 4 presents simulation results obtained from MATLAB 7.12.0, followed by concluding remarks in section 5.

2. Overview of transform techniques

In this paper DWT and DCT transforms are exploited for speech compression.

2.1. Discrete wavelet transform

A wavelet is a waveform of effectively limited duration that has an average value zero. Fundamental idea behind wavelet is to analyse according to scale[4]. Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. wavelet functions are localized in space. In contrast Fourier sine and cosine functions are non-local and are active for all time t. This localization feature, along with wavelets localization of frequency, makes many functions and operators using wavelets "sparse", when transformed into the wavelet domain. This sparseness, in turn results in a number of useful applications such as data compression, detecting features in images and de -noising signals. The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two , so called dyadic scales and positions. The mother wavelet is rescaled or dilated, by powers of two and translated by integers.Specifically, a function $f(t) \in L^2(\mathbf{R})$ (defines space of square integrable functions) can be represented as[9].

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\psi(2^{-j}t-k) + \sum_{k=-\infty}^{\infty} a(L,K)\phi(2^{-L}t-k)$$

The function ψ(t) is known as the mother wavelet, while φ(t) is known as the scaling function. The set of functions

$$\left\{ \sqrt{2^{-L}}\phi(2^{-L}t-k); \sqrt{2^{-j}}\psi(2^{-j}t-k) \mid j \le L, j,k,L \in Z \right\}$$

where Z is the set of integers, is an orthonormal basis for L2(**R**).

The numbers $a(L,k)$ are known as the approximation coefficients at scale L, while $d(j,k)$ are known as the detail coefficients at scale j.

The approximation and detail coefficients can be expressed as:

$$a(L,k) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{+\infty} f(t)\phi(2^{-L}t-k)dt$$

$$d(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} f(t)\psi(2^{-j}t-k)dt$$

wavelets with a high number of vanishing moments lead to a more compact signal representation and are hence useful in coding applications[9].

2.2 Discrete cosine transform

We can use DCT for speech compression because of high correlation in adjacent coefficient. Energy compaction property of DCT is good, we can often reconstruct a sequence very accurately from only a few DCT coefficients, this property of DCT is very useful for applications requiring data reduction[7].

Discrete cosine transform of 1-D sequence x(n) of length N is

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

Where m=0,1,…..,N-1.

The inverse discrete cosine transform is defined as

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

In both equations c_m is defined as

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

$\qquad 1 \qquad$ for m≠0

3. METHODOLOGY FOR COMPRESSION OF SPEECH SIGNAL

In this paper we are implementing speech compression technique based on transform method i.e. DCT and DWT. When we apply wavelet transform technique on speech signal original Signal can be represented in terms of a wavelet expansion (using coefficients in a linear combination of wavelet function),similarly in case of DCT transform speech can be represented in terms of DCT coefficient. Thus data operation can be performed using just the corresponding DWT and DCT coefficients. Transform techniques and thresholding does not actually compress a signal, it simply provides information about the signal, which allows the data to be compressed by standard encoding techniques.

Speech compression is achieved by neglecting small coefficients as insignificant data and discarding them and then applying quantization and encoding scheme on coefficients. Speech compression algorithm is performed in following steps:

(1). Transform technique on speech signal
(2). Thresholding of transformed coefficient.
(3). Quantization
(4). Encoding

By following above 4 steps we will get a compressed speech signal. For reconstruction of speech signal inverse of above processes are performed i.e. decoding, de-quantization, inverse

transform. 4 different speech signal samples are used for data base. Data base used are in .wav format, sampled at 8khz.

Explanation for compression algorithm:

3.1. Transform technique on speech signal

DCT and DWT transform methods are used on speech signal. We can reconstruct a sequence very accurately from only a few DCT coefficient, this property of DCT transform is used for data compression. Localization feature of wavelet along with time-frequency resolution property makes them well suited for speech coding. Sparse coding of wavelet is utilized for compression application. The idea behind signal compression using wavelets is primarily linked to the relative scarceness of the wavelet domain representation for the signal. Wavelets concentrate speech information (energy and perception) into a few neighbouring coefficients [8].
Therefore as a result of taking the wavelet transform of a signal, many coefficients will either be zero or have negligible magnitudes. Data compression is then achieved by treating small valued coefficients as insignificant data and thus discarding them.

3.2. Thresholding

After getting coefficient signal from different transform methods thresholding is applied to signal. In case of DCT as we mentioned earlier that very few DCT coefficient represent 99% of signal energy, hence threshold based on above mentioned property is calculated and applied to coefficient obtained from DCT transform, coefficients of which values are less than thresholds value are truncated ie. set to zero. In case of DWT we are applying level dependent thresholding threshold value is obtained from Birge-Massart strategy [9]. If we apply global thresholding, threshold value will be set manually [1].

3.3. Quantization

It is a process of mapping a set of continuously valued input data to a set of discrete valued output data. Aim of quantization step is to decrease the information found in threshold coefficients, in such a way that quantization process produce no error. We are performing uniform quantization process. For quantization maximum and minimum value of truncated coefficients m_{max} and m_{min} respectively is determined, and number of quantization level, L is

selected. Step size Δ is obtained using above 3 parameters.

$$\Delta=(m_{max}-m_{min})/L$$

Input is divided into L+1 levels with equal interval size ranging from m_{min} to m_{max} to create the quantization table. Coefficient values are quantized to integer values[1].
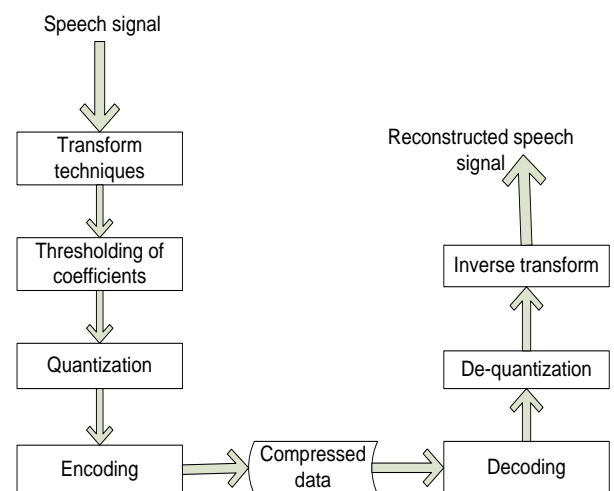

Fig. 1. Block diagram of compression technique

3.4. Encoding
In this paper we are using 3 different approaches for encoding of quantized coefficients:

1) Run length encoding: Coefficients obtained after Quantization contains many runs ie. same data value occurs in many consecutive data elements. Hence Run length encoding can be efficiently used for encoding of coefficients. Run length encoding is lossless data compression technique in which runs of data is stored as a single data value and count. In this paper our approach is to store data value and count in two different vectors. This encoding technique is applied on quantized coefficient obtained from both (DCT and DWT) transform technique.

2) *Huffman encoding:* Coefficients obtained after quantization process contains some redundant data ie. repeated data. Hence we can apply Huffman encoding. For Huffman encoding. Huffman encoding requires statistical information about data being encoded. For statistical information probabilities of occurrences of the symbol in quantized coefficient is computed,

symbols are arranged in ascending order along with their probabilities. The main computational step in encoding data from this source using a Huffman code is to create a dictionary that associates each data symbol with a codeword [1].

Run length encoding followed by Huffman encoding: Our 3rd approach is to perform run length encoding on truncated and quantized coefficient, and apply Huffman encoding on symbols (data values) and their counts separately. For reconstruction of speech signal inverse process of above mentioned processes i.e. decoding,de-quantization and inverse transform is performed.

4. Performance Measures

Performance of speech compression techniques based on 2 different transform methods is measured for 3 encoding approaches. Here performance is measured in terms of CR, SNR ,PSNR, NRMSE, and RSE.

1) Compression Factor:

CF= Length of original signal
          Length of compressed signal

for the compressed signal we have to take into account all the values that would be needed to completely represent the signal.

2) Signal to noise ratio:

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

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:

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

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.

4) Normalized root mean square error:

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

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

5) Retained signal energy:

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

$\|(x`(n)\|^2$ and $\|(x(n)\|^2$ represent energy of reconstructed and original speech signal. RSE indicates the amount of energy retained in the compressed signal as a percentage of the energy of original signal.

5. SIMULATION RESULTS

Experiments are conducted on 4 different sentences. Here results are obtained using different transform techniques DCT and DWT(db10) and for both case 3 different encoding techniques are exploited. Performance measures obtained from all cases are compared. From table 1 we can see that average Compression Factor obtained from DCT with Run length encoding is 2.05 , in case of DCT with Huffman encoding is 5.56 ,and for DCT with Run length encoding followed by Huffman encoding CF is 5.95. For DWT average Compression Factor is 14.6938 with Run length encoding,

11.432 With Huffman encoding and 29.11 with Run length followed by Huffman encoding. Results shows that we can achieve average compression factor 29.11 when we are implementing speech compression technique using DWT with run length followed by Huffman encoding. Since coefficients obtained from DCT contains very small runs hence average compression factor in that case has minimum value 2.05. Other performance measures SNR,PSNR,NRMSE and RSE depends on Transform technique but not on encoding technique.

TABLE I

COMPARISON OF PERFORMANCE BASED ON DIFFERENT TRANSFORM AND DIFFERENT
ENCODING METHODS

Transform Technique	Speech signal	CF			SNR			PSNR		
		RLE	Huff	RLE+ Huff	RLE	Huff	RLE+ Huff	RLE	Huff	RLE+ Huff
DWT	S4.wav	12.9249	11.0380	25.7960	1.7525	1.7525	1.7525	20.2607	20.2607	20.2607
	S6.wav	12.3991	10.7182	24.5242	2.1581	2.1581	2.1581	2.1581	23.0731	23.0731
	SP10.wav	12.9249	11.6080	28.3600	1.2334	1.2334	1.2334	13.9402	13.9402	13.9402
	SP15.wav	20.5263	12.362	37.7662	1.3594	1.3594	1.3594	8.3044	8.3044	8.3044
DCT	S4.wav	2.3361	6.1646	7.0613	7.8735	7.8735	7.8735	32.1771	32.1771	32.1771
	S6.wav	2.0868	5.5213	5.7854	8.0589	8.0589	8.0589	33.6532	33.6532	33.6532
	SP10.wav	1.6095	4.6519	4.6111	8.1561	8.1561	8.1561	27.6174	27.6174	27.6174
	SP15.wav	2.1783	5.8974	6.3454	7.8168	7.8168	7.8168	27.2869	27.2869	27.2869

Transform Technique	Speech signal	RSE			NRMSE		
		RLE	Huff	RLE+ Huff	RLE	Huff	RLE+ Huff
DWT	S4.wav	65.9925	65.9925	65.9925	0.5847	0.5847	0.5847
	S6.wav	73.1896	73.1896	73.1896	0.5204	0.5204	0.5204
	SP10.wav	50.3752	50.3752	50.3752	0.7060	0.7060	0.7060
	SP15.wav	48.5701	48.5701	48.5701	0.7180	0.7180	0.7180
DCT	S4.wav	98.2396	98.2396	98.2396	0.1483	0.1483	0.1483
	S6.wav	98.4133	98.4133	98.4133	0.1539	0.1539	0.1539
	SP10.wav	98.1177	98.1177	98.1177	0.1462	0.1462	0.1462
	SP15.wav	98.5682	98.5682	98.5682	0.1611	0.1611	0.1611

6. **CONCLUSION**

We have concluded from experiments that when in transform based speech compression system we exploit DWT with run length followed by Huffman encoding we can represent speech signal with minimum data values, and reconstructed speech obtained is intelligible hence it is an effective technique for speech compression.

7. REFERENCES

[1] G. Rajesh, A. kumar and k. Ranjeet, *speech compression using different transform techniques,* IEEE 2011

[2] MichelMisiti,YvesMisiti,Georges*Oppenheim,Jean-Michel Poggi,wavelet toolbox for use with matlab.*

[3] Najih A. M,,Ramli A. R,Syed A.R., *comparing speech compression using wavelets with other speech compression technique,* IEEE 2003.

[4]   Amara Graps, *An introduction to wavelet*

[5]   Jalal Karam, and RaedSaad ,*The Effect of Different Compression Schemes on Speech Signals ,* WASET 2006

[6]   VarunSetia, Vinod Kumar,*Coding of DWT Coefficients using Run-length coding and Huffman Coding for the purpose of Color Image Compression,*International Journal of Computer and Communication Engineering 6 2012

[7].*www.mathworks.com/**help**/toolbox/signal/ref/**dct**.html* Cached

[8]W. Kinsner and A. Langi, .Speech and Image Signal Compression withWavelets,. *IEEE Wescanex Conference Proceedings*, IEEE, New York,   NY,1993, pp. 368-375.

[9]   Rao N.,ELEC 4801 THESIS PROJECT