Speech Compression using Discrete Wavelet Transform with Optimum Differential Pulse Code Modulation (DPCM)

Fadel S. Hassen

Received on :14 /1 / 2007
Accepted on: 30 / 6 /2008

Abstract

In this paper the performance of speech compression system using discrete wavelet transform (DWT) is investigated. Two methods are used for this purpose; in the first only discrete wavelet transform is used. In the second method linear predictive coding (LPC) in wavelet transform is used. Both methods are used differential pulse code modulation (DPCM) for quantization with optimum selection of quantization parameters using Lloyd algorithm. Huffman code is used for both methods to increase the compression ratio without effect on the quality performance of speech signal. The results show that DPCM with DWT gives the best performance than when it is used with LPC and DWT. For speech compression using DWT, for n=3 and 4, Db10 gives high quality measure and less CF. For n=1 and 2, Db2 gives better results for compression factor and quality measures. For speech compression using LPC and DWT, Db6 gives good performance.

Key words: Speech compression, discrete wavelet transform (DWT), LPC, DPCM and Huffman code

الخلاصة:

في هذا البحث تم تمثيل منظومة ضغط الكلام بأستخدام تحويل المويجةالمقطعة(DWT). طريقتين تم استخدامهما لهذا الغرض,في الطريقة الاولى استخدم فقط تحويل المويجة المقطعة. اما في الطريقة الثانية فقد تم استخدام شفرة التخمين الخطية (LPC) في مجال تحويل المويجة. في كلا الطريقتين تم استخدام (DPCM) لتكميم المعاملات مع أفضل أختيار لهذه المعاملات بأستخدام طريقة (Lloyd) . تم أستخدام شفرة (Huffman) لزيادة نسبة الضغط بدون التأثير على نوعية الكلام. اثبتت النتائج ان أستخدام DPCM مع DWT اعطى نتائج افضل من استخدامه مع LPC وDWT. لضغط الكلام بأستخدام DWT (n=3,4) , Db10 اعطت نوعية عالية ولكنCF قليل, بينما Db 2 (n=1,2) اعطت نتائج جيدة لعامل الضغط ومقياس النوعية. لضغط الكلام بأستخدام LPC و DWT , Db6 اعطت تمثيل جيد.

Elect. Eng. Dept College of Eng. Al-Mustansirya University

1-Introduction

Speech coding has been and still a major issue in the area of digital speech processing. Speech coding is the act of transforming the speech signal at hand, to a more compact form, which can then be transmitted with a considerably smaller memory. The motivation behind this is the fact that access to unlimited amount of bandwidth is not possible. Therefore, there is need to code and compress speech signals [1]. Today applications of speech coding and compression have become very numerous. Many applications involve the real time coding of speech signals, for use in mobile satellite communictions, cellular telephony, and audio for videophones or video teleconferencing systems. Other applications include the storage of speech for speech synthesis and playback, or for the transmission of voice at a later time. Some examples include voice mail systems, voice memo wristwatches, voice logging recorders and interactive PC software [2].

There are many contributions have been achieved in the field of speech compressoion. In [2] a wavelet based speech coder is implemented. The performance of the wavelet compression scheme on both male and female spoken sentences is compared. On male spoken sentence the scheme reaches a signal to noise ratio of 17.45 dB and compression ratio of 3.88, using a level dependent thresholding approach. In [3], Musab T. S. Al-Deen presents an algorithm to compress speech signals by using residual excited linear prediction (RELP) with wavelet transform. In this project, Minimum bit rate is 9.25 kbps for SNR=8.3 dB. In [4] a code excited linear predictive (CELP) with various quantization methods (such as scalar, vector and DPCM) are used to compress speech signal. In [5] adaptive packet wavelet transform and Psychoacoustic Modeling are combined together to compress audio signal.

In this paper speech compression using discrete wavelets transforms (DWT), LPC, optimum DPCM and Huffman coded are investigated.

2- Wavelet Thresholding

Thresholding operations are applied to the detail coefficients of the wavelet transform. In this paper two types of threshold are introduced:

1-Hard thresholding

Hard thresholding also called (kill/keep) strategy, which is simplest method and can be stated mathematically as [7]:

$$THR(Y)=\begin{cases} Y & |Y|\ge\lambda \\ 0 & |Y|<\lambda \end{cases}$$

(1) where THR(Y) represents the output value after thresholding the wavelet coefficients and λ is thresholding value.

2-Soft-thresholding

Soft-thresholding is an alternative scheme of hard thresholding and can be stated mathematically as [7]:

$$THR(Y)=\begin{cases} \mathrm{sign}(Y)(|Y|-\lambda) & |Y|\ge\lambda \\ 0 & |Y|<\lambda \end{cases}$$

(2) In this paper, the thresholding values (λ) are level dependent thresholds that are calculated using Birge-Massart strategy [2]. This thresholding scheme based on an approximation results from Birge and

Massart and is well suited for signal compression. This strategy keeps all of the approximation coefficients at the level of decomposition J. The thresholding values at level i starting from 1 to J are given by the formula:

$$\lambda_i = \frac{2 * \text{length of the coarsest approximation coefficients}}{(J+2-i)^{1.5}} \tag{3}$$

3- Speech Compression Techniques

3-1 Speech Compression Technique using Discrete Wavelet Transform (DWT)

Figure (1) shows the block diagram of speech encoder/decoder. Combined usage of discrete wavelet transforms, differential pulse code modulation (DPCM) and Huffman code are involved. In this method, The speech signal is transformed by wavelet analysis and quantized by DPCM then encoded using Huffman code to increase the compression ratio. When the approximation part of the wavelet coefficients is transmitted most of the energy is found, therefore, approximation of the wavelet coefficients is kept without thresholding while the detail of the wavelet coefficients is thresholded using level dependent threshold (see Equation (3)). The approximation and detail of the wavelet coefficients are quantized together using DPCM. Huffman code is used to increase the compression ratio without effect on the resolution of the speech signal.

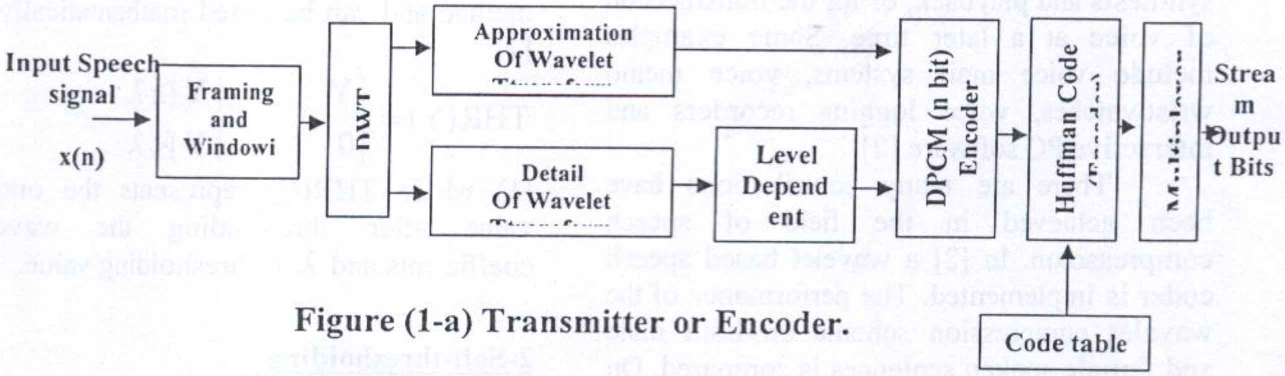

Figure (1-a) Transmitter or Encoder.

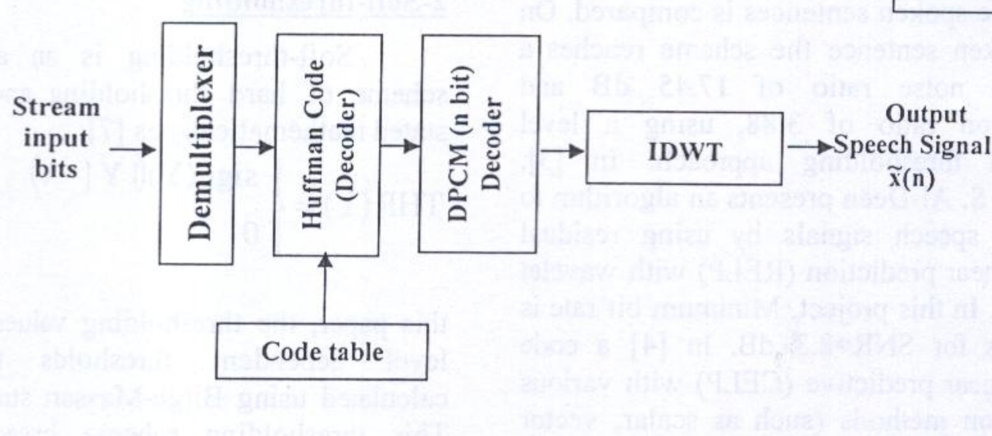

Figure (1-b) Receiver or Decoder

Figure (1) Speech Compression System using DWT

3-2 Speech Compression Technique using Linear Predictive Coding (LPC) and Discrete Wavelet Transform (DWT)

Figure (2) shows the block diagram of speech encoder/decoder. Combined usage of residual excited linear prediction prediction, wavelet transforms and differential pulse code modulation (DPCM) are involved. In this method, the LPC analysis provides the predictor coefficients (a_x) and the prediction residual error (e_n). The prediction residual error is transformed by wavelet analysis and quantized by DPCM. Befor quantized the residual error, the approximation coefficients are truncated. To ensure stability of the coefficients (the poles and zeros must lie within the unite circle in the z-plane) a relatively high accuracy (8 bit per coefficients is taken here) is required. This comes from the effect that small changes in the predictor coefficients lead to relatively large changes in the pole positions.

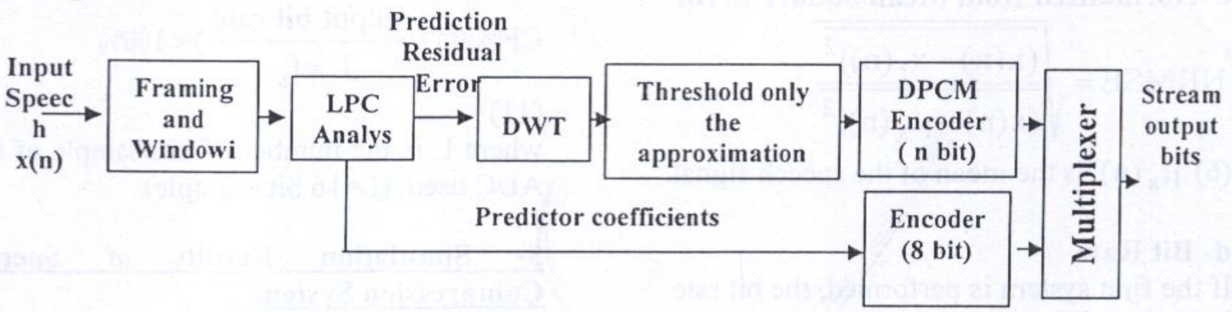

Figure (2-a) Transmitter or Encoder.

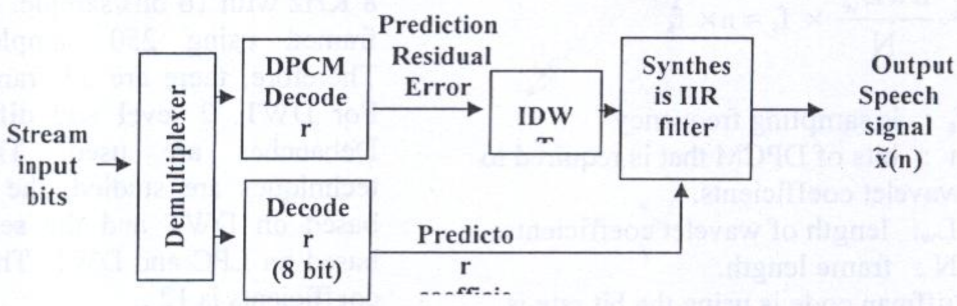

Figure (2-b) Receiver or Decoder

Figure (2) Speech Compression System using LPC and DWT

4- Performance Measures

Number of quantitative parameters can be used to evaluate the performance of speech coder in terms of both reconstructed signal quality after decoding and compression scores. The following parameters are used [2].

a- Signal to Noise Ratio (SNR)

This is defined perframe as:

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_{n=0}^{N-1} x_r^2(n)}{\sum_{n=0}^{N-1}\left(x(n)-x_r(n)\right)^2}\ \mathrm{dB} \quad (4)$$

where N: frame size.

$x(n)$: actual speech.

$x_r(n)$: reconstructed speech.

b-Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\left[\frac{NX^2}{\|x(n)-x_r(n)\|^2}\right] \text{dB}$$

(5) X is the maximum absolute value of the signal x, and

$\|x(n)-x_r(n)\|^2$ is the energy of the difference between the original and reconstructed signal.

c- Normalized Root Mean Square Error

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

(6) $\mu_x(n)$ is the mean of the speech signal.

d- Bit Rate

If the first system is performed, the bit rate is given by:

$$\text{Bit rate}=\frac{n\times L_w}{N}\times f_s \approx n\times f_s$$

(7)

where f_s : is sampling frequency

n : bits of DPCM that is required to encode wavelet coefficients.

L_w: length of wavelet coefficients.

N : frame length.

When Huffman code is using the bit rate is given by:

$$\text{Bit rate}=\frac{\ell_{ave}\times L_w}{N}\times f_s \approx \ell_{ave}\times f_s$$

(8) where ℓ_{ave} is the average code length of Huffman code.

If the second system is used, the bit rate is given by:

$$\text{Bit rate}=\frac{n\times N+p\times m}{N}\times f_s$$

(9)

where n : bits of DPCM that is required to encode residual error

b-Peak Signal to Noise Ratio (PSNR)

p : number of predictor coefficients.

m : bits required to encode predictor coefficients (here equals 8 bit)

When Huffman code is using the bit rate is given by:

$$\text{Bit rate}=\frac{\ell_{ave}\times N+p\times m}{N}\times f_s$$

(10)

e- Compression Factor (CF)

The compression factor is given by [6]:

$$CF\%=(1-\frac{\text{output bit rate}}{L\times f_s})\times 100\%$$

(11)

where L is the number of bits/sample of the ADC used. (L=16 bits/sample).

5- Simulation Results of Speech Compression System

A test speech message is sampled at 8 KHz with 16 bits/sample. This message is framed using 250 samples per frame. Therefore, there are 32 frames per second. For DWT, 2 level and different types of Debauches are used. There are two techniques are studied, the first technique based on DWT and the second technique based on LPC and DWT. The order of LPC coefficients is 12 .

5-1 Results of Speech Compression using DWT

Tables (1) ,(2), (3) and (4) show the performance quality measures of the reconstructed speech signals for different types of wavelets using n=4, 3, 2, and 1 respectively when hard threshold is used. From these tables it is shown that decreased number of bits of DPCM (n) reduced the performance of quality measure (SNR, PSNR and NRMSE) but increased the compression factor (CF). Also it is seen that increase the length of debauches is not

Db6) but increase the length of debauches decrease the compression factor.

Table (1) The performance measures concerning speech quality using different types of wavelet for n=4.

wavelet	SNR (dB)	PSNR (dB)	NRMSE	CF%	CF% using Huffman
Db2	20.5197	115.9510	0.1544	74.6	82.1845
Db4	22.1672	115.8160	0.1568	73.8	81.6640
Db6	23.4962	115.8101	0.1569	73.0	81.3374
Db8	20.6302	115.7425	0.1581	72.2	80.6299
Db10	26.6374	115.8580	0.156	71.4	80.0569

Table (2) The performance measures concerning speech quality using different types of wavelet for n=3.

wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
Db2	12.9714	111.944	0.2449	80.95	84.207
Db4	13.5928	111.670	0.2527	80.35	83.898
Db6	16.3339	111.942	0.2449	79.75	83.536
Db8	13.7785	111.421	0.2601	79.15	82.941
Db10	14.4306	111.541	0.2565	78.55	82.459

Table (3) The performance measures concerning speech quality using different types of wavelet for n=2.

wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
Db2	8.0653	107.7723	0.3958	87.3	88.125
Db4	7.5776	107.3038	0.4178	86.9	87.800
Db6	7.4916	107.1990	0.5370	86.5	87.493
Db8	7.2613	107.3025	0.4177	86.1	87.045
Db10	7.3887	107.1384	0.4258	85.7	86.689

Table (4) The performance measures concerning speech quality using different types of wavelet for n=1.

wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%
Db2	3.5000	103.4415	0.6517	93.65
Db4	3.3775	103.2034	0.6698	93.45
Db6	3.2378	103.1667	0.6727	93.25
Db8	3.2792	103.1480	0.6741	93.05
Db10	3.1617	103.0330	0.6831	92.55

Table (5) shows the bit rate for Db2 and n=4, 3, 2 and 1 for with and without Huffman code. From this table it is seen that the bit rate proportional inversely with CF and Huffman code appears good effect on bit rate at n increased more than 3 bit.

Table (5) the bit rate for Db2 and different types of n

	n=1	n=2	n=3	n=4
Bit rate (kbps) without Huffman	**8.1**	**16.2**	**24.3**	**32.5**
Bit rate (kbps) with Huffman code	**8.1**	**15.2**	**20.2**	**22.8**

Table (6) shows the performance quality measures of the reconstructed speech signals for different values of (n) when soft threshold and Db2 are used. It is seen that increase n will enhance the quality of speech signal but effect more badly on CF. therefore there are trade off between selections of n , SNR and CF.

Table (6) The performance measures concerning speech quality using soft thresholding and different values of (n) for Db2.

N	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
1	2.7145	103.2668	0.665	93.65	93.6500
2	5.8059	107.4681	0.555	87.30	88.2989
3	8.0010	111.2160	0.266	80.95	84.3733
4	9.2266	114.1727	0.189	74.60	82.5693

5-2 Results of Speech Compression using LPC and DWT

Tables (7), (8), (9), (10) and (11) show the performance quality measures of the reconstructed speech signals for different types of wavelets and (n) when hard threshold is used. From these tables it is seen that decreased n will increased CF but decreased quality measure (SNR, PSNR and NRMSE). For n=1 this is the highest CF can be obtained and no need to use Huffman code. Generally effect of Huffman code appears clearly at high number of bits (n more than 3 bits).

Table (7) The performance measures concerning speech quality using different types of wavelet for n=4.

Wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
Db2	5.2575	103.861	0.6211	72.1680	84.2193
Db4	4.3619	103.597	0.6403	73.8010	85.2088
Db6	4.9182	103.762	0.6282	73.0915	84.5029
Db8	4.3972	103.275	0.6288	73.0023	84.4491
Db10	4.3990	103.603	0.6390	71.0271	83.7869

Table (8) The performance measures concerning speech quality using different types of wavelet for n=3.

Wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
Db2	4.4233	103.6447	0.6368	78.540	84.9033
Db4	4.0275	103.4563	0.6508	79.764	86.1780
Db6	4.8704	103.4927	0.6480	79.23	85.5707
Db8	4.1779	103.6052	0.6397	79.165	85.5815
Db10	4.6370	103.3205	0.6588	77.684	84.8033

Table (9) The performance measures concerning speech quality using different types of wavelet for n=2.

Wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%	CF% using Huffman
Db2	3.2972	102.0990	0.7608	84.9121	87.9036
Db4	3.4070	102.4521	0.7305	85.7286	88.6868
Db6	3.6318	102.6626	0.7130	85.3739	88.1756
Db8	2.7937	102.3974	0.7351	85.3293	88.2466
Db10	3.0790	102.5062	0.7260	84.6417	87.5124

Table (10) The performance measures concerning speech quality using different types of wavelet for n=1.

Wavelet	SNR(dB)	PSNR(dB)	NRMSE	CF%
Db2	1.4627	99.2229	1.0595	91.2842
Db4	1.3242	99.1702	1.0659	91.6924
Db6	1.5847	99.4755	1.0291	91.5151
Db8	1.4349	99.4929	1.0271	91.4928
Db10	1.3078	99.7280	0.9996	90.9990

Table (11) shows the bit rate for Db6 with and without Huffman code. It can seen that minimum bit rate occurs at n=1 (11.1 kbps) with SNR=1.584 dB while for previous speech compression using DWT, minimum bit rate occurs at n=1 (8.1 kbps) with SNR=3.5 dB. Therefore, the performance of the first system is better than the second system.

Table (11) The bit rate for Db6 and different types of n.

	n=1	N=2	n=3	n=4
Bit rate (kbps) without Huffman	11.1	18.7	26.5	34.4
Bit rate (kbps) with Huffman code	11.1	15.1	18.4	19.8

6-Conclusions

The following are a summary of the conclusion remarks.

1. The tradeoffs between quality on one side of bit rate and complexity on the other side clearly appear here. If we want a better quality, the complexity of the system should be increased or a larger bit rate has to be used.
2. Thresholding value that is used in this paper gives good estimation about the redundancy of the coefficients that is eliminated it and don't effect on the performance of quality measure and hence thresholds this unimportant coefficients add other increasing for compression factor.
3. Hard thresholding gives the best results than soft thresholding.
4. For speech compression using DWT, for n=3 and 4, Db10 gives high quality measure and less CF. For n=1 and 2, Db2 gives better results for compression factor and quality measures.
5. For speech compression using LPC and DWT, Db6 gives good performance.
6. Huffman code gives variable compression factor for each frame and gives good increase for compression factor without effect on the performance of quality of speech signal. (Huffman code is not used with one bit quantization).
7. The results show that DPCM with DWT gives the best performance than when it is used with LPC and DWT.

Reference

1. Ozgu Ozun, Philipp Steurer and Daniel Thell, "**Wideband Speech Coding with Linear Predictive Coding (LPC)**", MSc. Thesis, California University, Winter 2002.
2. Nikhil Rao, "**Speech Compression Using Wavelet**", MSc. Thesis, School of Information Technology and Electrical Engineering, October 2001.
3. Musab T. S. Al-Deen, "**Speech Compression using Linear Prediction Coding in Wavelet Domain**", MSc. Thesis, Mustansiriya University, 2004.
4. Awis M. Kamboh, Krispian C. Lawerence, Aditya M. Thomas and Philips I. Tsai, "**Design of a CELP Coder and Analysis of Various Quantization T**
5. **echniques**", EECS 651 Project Report, University of Michigan, Ann Arbor, Winter 2005.
6. Matias Zanartu "**Audio Compression using Wavelet Techniques**", ECE648 Project Report, Electrical and Computer Engineering, Purdue University, Spring 2005.
7. Raghuveer M. Rao and Ajit S. Bopardikor, "**Wavelet Transform: Introduction to Theory and Application**", Addison-Wesley Longman, Inc., California, 1998.
8. Ahmed Kamil Hassan, "**Speech Denoising Using Wavelet Transform**", MSc. Thesis, Mustansiriya University, 2005.
9. Davis Yen Pan, "**Digital Audio Compression**", Digital Technical Journal Vol. 5, No.2, Spring 1993.
10. Lloyd, S. P., "**Least Squares Quantization in PCM**", IEEE Transactions
11. on Information Theory, Vol IT-28, March, 1982, pp. 129-137.