Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering
Shaw Conference Center, Edmonton, Alberta, Canada May 9-12 1999

COMPARING AUDIO COMPRESSION USING WAVELETS WITH OTHER AUDIO COMPRESSION SCHEMES

El-Bahlul Fgee, W. J. Phillips and W. Robertson
Department of Electrical and Computer Engineering
Dalhousie University, DalTech
P.O. Box 1000
Halifax, Nova Scotia, Canada B3J 2X4

Abstract

Speech compression is the technology of converting human speech into an efficient encoded represented that can be decoded to produce a close approximation of the original signal [1]. In this paper, we propose a new algorithm which compresses speech signals using wavelet compression technique. The performance of this method is compared against the following representative coding and compression shemes [3]. Adaptive Differential Pulse Code Modulation (ADPCM) which reduces the transmitted data by a factor of two. Linear Predictive Coding (LPC) which uses compression ratio of more than twelve to one. Linear Predictive Coding algorithm using the United States Department of Defense Standard 1015 used compression ratio of 26:1. Global System Mobile (GSM) algorithm which reduces the transmitted data by a factor of five. The following parameters are compared :
- Quality of the reconstructed signal after decoding .
- Compression ratios .
- Signal to Noise Ratio (SNR).
- Peak Signal to Noise Ratio (PSNR).
- Normalized Root Mean Square Error (NRMSE).

INTRODUCTION

Compression of wide band audio signals to very low bit rates is desirable for a number of applications such as transmission of digital audio , multimedia applications [1]. Multimedia and video conferencing, dynamic Web-site access with voice and video introduces the idea of using voice over the Internet. This idea also opens up new commercial opportunities in the area of self-service and other service applicability [2]. In this paper we used wavelets to compress speech signals and compare the results with other compression methods [3]. We start the process by dividing the speech signal into segments (20 msec length) since 20 msec speech segments gives second order stationary in time. We apply different types of wavelets, Daubechies wavelets(db4, db6, db10, db20), and Biorthogonal (3.9) wavelets. We use these wavelets with both a global and a level dependent thresholding techniques at different compression ratios and at different decomposition levels to find the best wavelet model and the optimum thresholding technique. The best wavelet with the optimum thresholding technique is compared with the other compression schemes in the next step. As we know wavelets work by decomposing a signal into different resolutions or frequency bands [5], and this task is carried out by choosing the wavelet function and computing the Discrete Wavelet Transform (DWT). This analysis is carried in two main process, wavelet decomposition and wavelet reconstruction. The decomposition procedure (Analysis filter bank) breaks the signal into pieces and the reconstruction procedure (Synthesis filter bank) puts these pieceses back together again [6].These processes is done by filtering and downsampling (in decomposition procedure) and filtering and upsampling (in reconstruction procedure) [5].
Finally, the best results obtained from this wavelet model and this thresholding technique are compared with other compression schemes [3] in terms of the previous measuring parameters.

SPEECH COMPRESSION

The goal of speech compression is to represent a signal using the smallest number of data bits commensurate with acceptable reconstruction. Wavelets concentrate speech information (energy and perception) into a few neighbouring coefficients [4]. This

0-7803-5579-2/99/$10.00 © 1999 IEEE

means a small number of coefficients will remain and others will be truncated. After decomposing a signal we apply a threshold to coefficients for each level from level 1 to level N (last decomposition level) [5]. The threshold is selected and a hard threshold is applied to coefficients by using a global thresholding [7]. The signal is reconstructed from the remaining coefficients.

Global Thresholding :-

Global thresholding works by retaining the wavelet transform coefficients which have the largest absolute value. This algorithm starts by dividing the speech signal into frames of equal size F. The wavelet transform of a frame has a length T (larger than F). These coefficients are sorted in descending order and the largest L coefficients are retained. In any application these coefficients along with their positions in the wavelet transform vector must be stored or transmitted. That is, $2L$ coefficients are used instead of the original F samples. The compression ratio,C, is therefore:

$$C = \frac{F}{2L} \quad or \quad L = \frac{F}{2C} \tag{1}$$

Each frame is reconstructed by replacing the missing coefficients by zeros. The above procedures are shown in Figure (1).

OTHER SCHEMES

As we know speech compression is the technology of converting human speech into an efficient digital encoding that can later be decoded to produce a close audio approximation of the original signal. In this work we have used the Speak Freely System [3] to implement the four different compression schemes with different compression ratios to compare them with the wavelet performance. We describe these schemes and give brief information about each one of them below.

1-ADPCM :-

ADPCM Compression uses Adaptive Differential Pulse Code Modulation to halve the data from 64 Kbps to 32 Kbps. The encoder consists of a quantizer and a linear predictor, and the decoder consists of a linear predictor . In this method both the quantizer and the predictor are adapted to the speech signal [8].

2-LPC :-

LPC compression uses Linear Predictive Coding to reduce the data rate by more than a factor of twelve, the basic idea behind the LPC model is that for a given speech samples at time n, s(n) can be approximated as a linear combination of the past p speech samples [9].

3-GSM :-

GSM Compression uses the algorithm GSM (Global System Mobile) which reduces the data rate by a factor of almost five with little degradation of voice-grade audio . This algorithm is based on a Regular Pulse Excited-Linear Predictive Coder (RPE-LPC) which compresses speech with two filters and an initial excitation [10][11].

4-LPC-10 :-

LPC-10 compression uses a different form of linear predictive coding, as specified by the United States of Defense as Federal Standard 1015/Nato-STANAG-4198 republished as a Federal Information Processing Standard 137 (FIPS Pub 137). LPC-10 compression encodes real-time audio into a 2400 bps stream compressing signal by a factor of more than 26 to 1 [12] [13].

SIMULATION and RESULTS

This section is divided into two subsections, in the first subsection we present the results obtained from trying different wavelets with both threshold techniques. In the second subsection we compare the results obtained from the best wavelet model (db10) with the other compression schemes previously mentioned [3].

Section I :-

The wavelets used in this section are Daubechies(db4, db6, db10, db20), and Biorthogonal (3.9) wavelets[14]. The results obtained for SNR, PSNR, NRMSE are calculated using the following formulas:

1- Signal to Noise Ratio:

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

where ${\sigma_x}^2$ is the mean square of the speech signal, ${\sigma_e}^2$ is the mean square difference between the original and reconstructed signals [15].

2- Peak Signal to Noise Ratio :

$$PSNR = 10\log_{10}\frac{NX^2}{||x - \tilde{x}||^2} \tag{3}$$

where N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x, and $||x - \tilde{x}||^2$ is the energy of the difference between the original and reconstructed signals respectively [16].

3- Normalized Root Mean Square Error :

$$NRMSE = \sqrt{\frac{\sum_n (x(n) - \tilde{x(n)})^2}{\mu_n (x(n) - \mu_x(n))^2}} \tag{4}$$

where $x(n)$ is the speech signal, $\tilde{x(n)}$ is the reconstructed signal, and $\mu_x(n)$ is the mean of the speech

signal [17].

Table I shows the average of the previous measuring parameters for the digit "two" taken form the TIDIGITS collection [18] and spoken by six males and six females. We used the global threshold technique with a frame size 160 samples (20 msec) at the fifth decomposition level. The compression ratio was fixed (26:1 and 12:1) and five different wavelets used. The compression ratios of 26:1 and 12:1 are chosen to allow direct comparison with the other compression schemes. As we can see from Table I the best wavelet is *db*10 since it gave the highest SNR, PSNR at the same level using the same thresholding technique. So, we will use this wavelet to compare it with the other compression techniques.

Section II :-

In this section we compare the quality of the reconstructed signal after compression using *db*10 wavelet and the global thresholding with the linear predictive coding (LPC), LPC-10, global system mobile (GSM), and adaptive differential pulse coding modulation (ADPCM) [3] in terms of the previous measuring parameters. The results obtained are shown in Table II. The waveforms of the original and reconstructed signals for the digit "two" spoken by males (in Figure (2a,b,c)) for compression ratios 26.67:1, 12:1, 5:1 using *db*10 wavelet and LPC-10, LPC, GSM. We have listened to the reconstructed signals after compression by these techniques, and we found that the wavelet gave better speech quality than LPC-10, LPC, but GSM and ADPCM gave almost the same results as the wavelet compression technique.

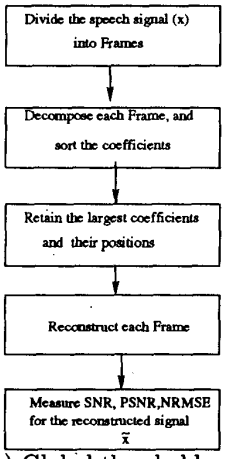
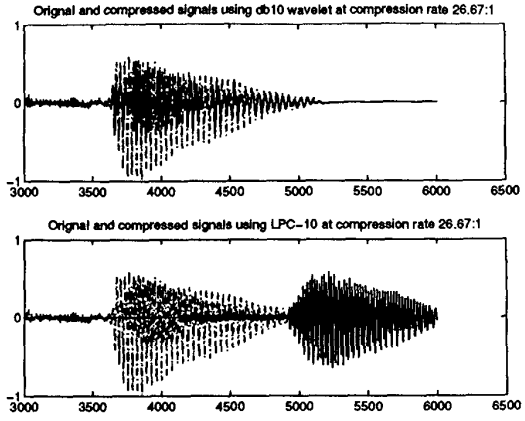


Figure (1) Global threshold procedures



Fig(2a) Original & Compressed signals using db10 and LPC-10 at 26.67:1

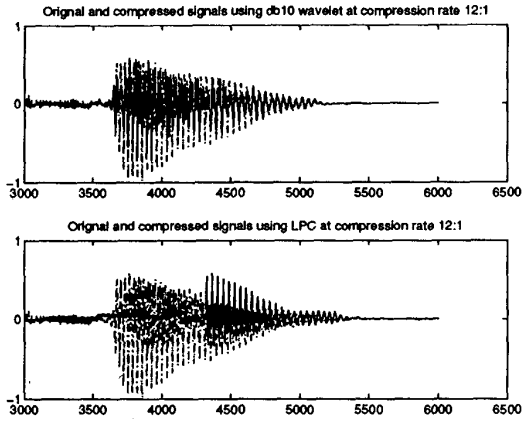


Figure (2b) Original and Compressed signals using db10 and LPC at 12:1

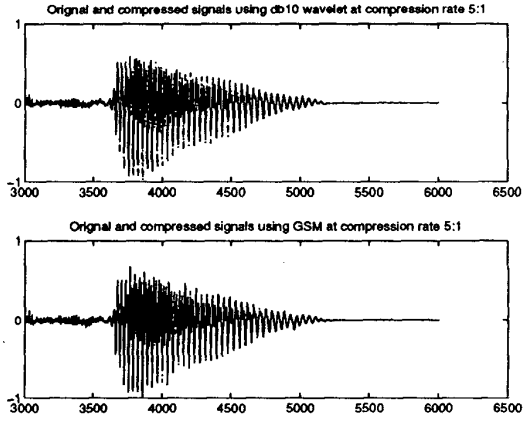


Figure (2c) Original and Compressed signals using db 10 and GSM at compression ratio 5:1 (for male speaker)

Wavelet Type	Comp. Ratio	SNR db		PSNR db		NRMSE	
		M	F	M	F	M	F
db4	26:1	2.73	2.16	19.47	19.02	0.73	0.78
	12:1	5.01	3.91	21.76	20.77	0.56	0.64
db6	26:1	2.79	2.30	19.53	19.16	0.73	0.77
	12:1	5.03	4.06	21.78	20.92	0.56	0.63
db10	26:1	2.98	2.27	19.74	19.12	0.71	0.77
	12:1	5.08	4.17	21.83	21.03	0.56	0.62
db20	26:1	2.65	2.48	19.39	19.34	0.74	0.75
	12:1	4.73	4.40	21.47	21.25	0.58	0.61
Bior (3.9)	26:1	1.68	0.73	18.43	17.56	0.83	0.92
	12:1	3.35	2.27	20.09	19.08	0.68	0.74

Table I Comparison between different wavelets using word two spoken by six Male and Female speakers at compression ratios 26:1 and 12:1 using a global threshold technique

Comp. Scheme	Comp. Ratio	SNR db		PSNR db		NRMSE	
		M	F	M	F	M	F
db10	26:1	2.98	2.27	19.74	19.12	0.71	0.77
LPC-10	26:1	-2.47	-2.46	14.32	14.31	1.34	1.35
db10	12:1	5.08	4.17	21.83	21.03	0.56	0.62
LPC	12:1	-1.81	-1.46	14.82	15.29	1.27	1.20
db10	5:1	10.27	8.55	27.48	25.41	0.31	0.37
GSM	5:1	14.89	14.51	31.63	31.36	0.19	0.19
db10	2:1	16.55	15.70	33.29	32.56	0.15	0.17
ADPCM	2:1	20.42	23.80	37.15	40.66	0.10	0.07

Table II Comparison between db10 and LPC-10, LPC, GSM, ADPCM in terms SNR, PSNR, NRMSE for digit "two" spoken by six male and female speakers

CONCLUSIONS

As a result from these experiments, using wavelets gives higher SNR and better speech quality than linear predictive coding (LPC, LPC-10), and comparable speech quality as the global system mobile (GSM) and adaptive differential pulse code modulation (ADPCM). In addition, using wavelets, the compression ratio can be easily varied while other compression schemes have fixed compression ratios. Finally, wavelets are less computionaly intensive than the other compression schemes.

References

[1] D.Sinha, J.D. Johnston. Audio compression at low bit rates using adaptive switched filter bank, Processing of IEEE on Acoustics, Speech and Signal Processing, Vol. 2 pp. 1053-1056, 1996.

[2] Daniel Minoli and Emma Minoli, *Delivering Voice over iP Networks*, Published by John Wiley & Sons, Inc. 1998.

[3] `http:\\www.fourmilab.ch\speak free`

[4] W.Kinser and A. Langi. Speech and Image signal compression with wavelets, IEEE, 1993.

[5] E. Mandridake, M. Najin. Joint wavelet transform and vector quantization for speech coding, IEEE, 1993.

[6] Randy K. Young, *Wavelet theory and its applications*, Kluwer academic Publishers 1995.

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

[8] A. Syral, R. Bennett and S. Greenspan, "Applied Speech Technology"

[9] L.Rabiner, B. Juang , Fundamental of Speech Recognition, Prentice Hall 1993.

[10] John Scourias . Overview of the Global System for Mobile Communication.

[11] Dr. Dobb's Web site, Jutta Degener, Digital Speech Compression (Putting the GSM 06.10 RPE-LTP algorithm to work, Dec.1994.

[12] Analog to Digital Conversion of Voice by 2400 bit/second Linear Predictive Coding, Federal Standard 1015, Nov 1984.

[13] T. E. Termain, The Government Standard Linear Predictive Coding Algorithm: LPC-10, Speech Technology, April 1982.

[14] M. Misiti, Y. Misiti, G. Oppenheim, J. Poggi. Matlab wavelet tool box, 1997.

[15] Lawrence R. Rabiner, Ronald W. Schafer , Digital Processing of Speech Signals, Prentice Hall 1978.

[16] Carl Taswell, Speech compression with cosine and wavelet packet near-best bases, Acoustic, Speech, and Signal Processing, 1996 ICASSP-96. Conference proceedings. 1996 Vol. 1.

[17] S. C. Sivakumar, W. Robertson, W. J. Phillips, "On-Line Stabilization of Block-Diagonal Recurrent Neural Networks", IEEE Transaction on Neural Networks, Vol. 10, No.1, Jan. 1999.

[18] NIST, Speech Discs, "Studio Quality Speaker-Independent Connected-Digital Corpus", NTIS PB91-506592 Texas Instruments, Feb. 1991.