UNIVERSITI PUTRA MALAYSIA

SPEECH COMPRESSION USING DISCRETE WAVELET TRANSFORM

ABDULMAWLA M. ALI NAJIH

SPEECH COMPRESSION USING DISCRETE WAVELET TRANSFORM

By

ABDULMAWLA M. ALI NAJIH

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Partial Requirements for the Degree of Master of Science

August 2003

In the name of God, Most Gracious, Most Merciful

Dedicated to,

My wife & my family

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the partial requirements for the degree of Master of Science

SPEECH COMPRESSION USING DISCRETE WAVELET TRANSFORM

By

ABDULMAWLA M. ALI NAJIH

August 2003

Chairman: Associate Professor Abdul Rahman Ramli, Ph.D.

Faculty: Engineering

Speech compression is an area of digital processing that is focusing on reducing bit rate of the speech signal for transmission or storage without significant loss of quality. Wavelet transform has been recently proposed for signal analysis. Speech signal compression using wavelet transform is given a considerable attention in this thesis.

Speech coding is a lossy scheme and is implemented here to compress one-dimensional speech signal. Basically, this scheme consists of four operations which are the transform, threshold techniques (by level and global threshold), quantization, and entropy encoding operations. The reconstruction of the compressed signal as well as the detailed steps needed are discussed.

The performance of wavelet compression is compared against linear Productive Coding and Global System for Mobile Communication (GSM) algorithms using SNR, PSNR, NRMSE and compression ratio.

Software simulating the lossy compression scheme is developed using Matlab 6. This software provides the basic speech analysis as well as the compression and decompression operations. The results obtained show reasonably high compression ratio and good signal quality.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi sebahagian keperluan untuk ijazah Master Sains

PEMAMPATAN SUARA MENGGUNAKAN PEMINDAHAN ANAK ISYARAT TERPUTUS

Oleh

ABDULMAWLA M. ALI NAJIH

Ogos 2003

Pengerusi: **Profesor Madya Abdul Rahman Ramli, Ph.D.**

Fakulti: **Kejuruteraan**

Pemampatan suara adalah suatu bidang pemprosesan isyarat digital yang memberikan pemerhatian terhadap pengurangan kadar bit isyarat untuk penghantaran atau penyimpanan tanpa mengurangkan kualiti suara. Baru-baru ini, teknik pemindahan anak isyarat telah dicadangkan untuk pemprosesan isyarat. Tesis ini memberikan perhatian yang khusus kepada pemampatan isyarat suara menggunakan anak isyarat.

Pengekodan suara adalah dengan menggunakan skema kehilangan dan dijalankan untuk memampatkan isyarat suara satu dimensi. Skema ini mengandungi empat kaedah operasi iaitu pemindahan, teknik-teknik ambang (peringkat dan global), kuantisasi dan entropi. Perbincangan mengenai pembinaan semula isyarat yang telah dimampatkan dan langkah-langkah yang diperlukan adalah dinyatakan.

Prestasi pemampatan menggunakan anak isyarat dibandingkan dengan algoritma-algoritma LPC serta GSM berdasarkan SNR, PSNR, NRMSE dan nisbah pemampatan.

Simulasi skema pemampatan kehilangan adalah dibangunkan dengan menggunakan perisian MATLAB versi 6. Perisian ini menyediakan analisis asas suara serta operasi pemampatan dan penyah-mampatan. Keputusan-keputusan yang didapati menunjukkan nisbah mampatan yang tinggi dan kualiti isyarat yang baik.

ACKNOWLEDGEMENTS

First of all, I would like to express my utmost thanks and gratitude to Almighty Allah S.W.T for giving me the ability to finish this thesis successfully.

The author gratefully with to express his profound appreciation and gratitude to his supervisor, Dr.Hj. Abdul Rahman Ramli , for his supervision , guidance, supporting, and constructive suggesting and comment throughout the duration of the project until it turn to real success

The authors also indebted to members of his supervisory committee, for their affectionate guidance, prompt decision and valuable assistance during this period.

Appreciation also to the assistance rendered by the respective lecturers, staffs, technicians of faculty of engineering for providing the facilities required for undertaking this project.

The author would like to thank his family for the encouragement and support without which is impossible for the success of this project, and my friends, especially Menchwy Kmees for offering helps all the time

LIST OF CONTENTS

		Page
DEDICATION		ii
ABSTRACT		iii
ABSTRAK		v
ACKNOWLEDGEMENTS		vii
APPROVAL		ix
DECLARATION		x
LIST OF TABLES		xiii
LIST OF FIGURES		xiv
CHAPTER		
I	INTRODUCTION	1
	1.1 Motivation for Studying this Problem	3
	1.2 Objectives	4
	1.3 Organization of the Thesis	5
II	LITERATURE REVIEW	6
	2.1 Human Speech Production	6
	2.2 Overview of Speech Coding	8
	2.2.1 Waveform Coders	9
	2.2.2 Source Coders	10
	2.2.3 Hybrid Coders	12
	2.3 Speech Coding Standards	14
	2.3.1 Pulse Code Modulation (PCM G.711)	15
	2.3.2 Adaptive Different Pulse Code Modulation (ADPCM, G.721, G.726, G.727)	16
	2.3.3 Linear Predictive Coding LPC	17
	2.3.4 Code Excited Linear Prediction Voice Coder CELP	17
	2.3.5 Global System for Mobile Communication GSM	18
	2.4 Previous work on wavelet speech coding	19
	2.5 Compression Techniques	22
	2.5.1 Lossless Compression	22
	2.5.2 Lossy Compression	22
	2.6 Wavelets Transform	23
	2.6.1 Wavelets vs. Fourier Analysis	24
	2.6.2 Time-Frequency Resolution	26
	2.7 Wavelets Families	28
	2.7.1 Haar Wavelet	29
	2.7.2 Daubechies Orthogonal Wavelets.	30
	2.7.3 Comparison of Wavelet Properties	33
	2.8 The Discrete Wavelet Transform	34
	2.8.1 Vanishing Moments	36
	2.9 The Fast Wavelet Transform Algorithm	37
	2.9.1 Implementation Using Filters	37
	2.9.2 Multilevel Decomposition	40
	2.9.3 Signal Reconstruction	41
	2.10 Thresholding	42

	Conclusion	43
III	**METHODOLOGY**	45
	3.1 Introduction	45
	3.2 Implementation stages	46
	3.2.1 Speech Signal Selection	47
	3.2.2 Discrete Wavelet Transform	47
	3.2.3 Choosing the threshold values	53
	3.2.4 Quantization	54
	3.2.5 Huffman Encoding	58
	3.2.6 Quality evaluation	62
	3.2.7 Speak Freely	64
IV	**RESULTS AND DISCUSSION**	65
	4.1 Analysis of speech using Wavelet Transform	65
	4.1.1 Selection of wavelet filter	66
	4.2 Speech Coding using independent level threshold	69
	4.2.1 The effect of decomposition levels on local threshold	71
	4.3 Speech Coding using global threshold	76
	4.3.1 the effect of changing decomposition levels on global threshold	78
	4.3.2 the effect of changing threshold values on compression Performance	81
	4.4 The Effect of Changing quantization level	83
	4.5 Comparative to other Schemes	84
V	**CONCLUSION AND RECOMMENDATION**	86
	5.1 Conclusion	86
	5.2 Recommendations for future work	88
REFERENCES		89
APPENDIX		94
VITA		101

LIST OF TABLES

Table		Page
2.1	Digital speech coding standards	15
2.2	Coefficients for the 4-tap Daubechies Low-pass Filter	31
2.3	Coefficients for the 6-tap Daubechies Low-pass Filter	31
2.4	Coefficients for the 10-tap Daubechies Low-pass Filter	32
2.5	Comparisons of two kinds of wavelets	33
4.1	Percentage of speech signal concentrated by wavelets in N/2 transform Coefficients	67
4.2	SNR, PSNR and NRMSE for Male speech signal	69
4.3	SNR, PSNR and NRMSE for female speech signal	70
4.4	Compression ratios for male and female speech signal	70
4.5	Different quantization levels versus NRMSE	76
4.6	SNR, PSNR and NRMSE for female speech signal	77
4.7	Compression ratios for male and female speech signals	77
4.8	Different quantization levels versus NRMSE	83
4.9	Comparative between Db10, LPC and GSM	85

LIST OF FIGURES

Figure **Page**

Figure		Page
2.1	Quality comparison of speech coding schemes	8
2.2	Generalised Block Diagram of a Vocoder a) Analysis at the transmitter b) Synthesis process at the receiver	11
2.3	Analysis-by-Synthesis, (AbS) coder structure. (a) Encoder (b) Decoder.	13
2.4	WFT Resolution	27
2.5	Wavelets Resolution	28
2.6	Haar Scaling Function	29
2.7	Haar Mother Wavelet	30
2.8	Daubechies 4-tap Scaling and Wavelet Functions	31
2.9	Daubechies 6-tap Scaling and Wavelet Functions	32
2.10	Daubechies 10-tap Scaling and Wavelet	33
2.11	Filtering operation of the DWT	39
2.12	Decomposition of DWT coefficients	40
2.13	Wavelets Reconstruction	42
3.1	Speech encoder and decoder	46
3.2	Analysis / resynthesis process for a three level	48
3.3	Flowchart for one-dimensional forward DWT: a) General Decomposition, b) Detail decomposition breakdown	51
3.4	Flowchart for one-dimensional inverse DWT: a) General Reconstruction, b) Detailed reconstruction breakdown	52

3.5 Flowchart for uniform quantization: (Quantizer) 55

3.6 Uniform quantizer 57

3.7 Flowchart for uniform quantization (Dequantizer) 58

3.8 Flow chart for Huffman encoding (Encoder) 60

3.9 Flow chart for Huffman encoding: (Decoder) 61

4.1 Speech signal , after one level decomposition 66

4.2 Compression ratios vs. wavelet filters 71

4.3 The effect of decomposition levels on local threshold 72

4.4 SNR versus decomposition levels for two-speech signal 72

4.5 Compression ratio versus decomposition levels for two-speech signal 73

4.6 Original Speech Signal and Reconstructed Approximat 75

4.7 ions at Different scales

4.7 Compression ratios vs. filters 78

4.8 The effect of decomposition levels on global threshold 79

4.9 SNR versus decomposition levels for two-speech signal 79

4.10 Compression ratio versus decomposition levels for two-speech signal 80

4.11 SNR vs. Threshold 81

4.12 PSNR vs. Threshold 82

4.13 NRMSE vs. Threshold 82

4.14 NRMSE versus quantization levels 84

CHAPTER I

INTRODUCTION

Speech coding has been and still is a major issue in the area of digital speech processing. Speech coding is the act of transforming the signal speech at hand, to a more compact form, which can then be transmitted with a considerably smaller size. The motivation behind this is the fact that access to unlimited amount of bandwidth is not possible. Therefore, there is a need to code and compress speech signals. Speech compression is required in long-distance communication, high-quality speech storage, and message encryption. For example, in digital cellular technology many users need to share the same frequency bandwidth. Utilizing speech compression makes it possible for more users to share the available system. Another example where speech compression is needed is in digital voice storage. For a fixed amount of available memory, compression makes it possible to store longer messages.

Speech coding is a lossy type of coding, which means that the output signal does not exactly sound like the input. The input and the output signal could be distinguished to be different. Coding of audio however, is a different kind of problem than speech coding. Audio coding tries to code the audio in a perceptually lossless way. This means that even though the input and output signals are not mathematically equivalent, the sound at the output is the same as the input. This type of coding is used in applications for audio storage, broadcasting, and Internet streaming (Zaki and Hosny 2001)

Traditionally speech coders can be classified into two categories: waveform coders and analysis-synthesis vocoders (from voice coders). Waveform coders attempt to copy the actual shape of the signal produced by the microphone and its associated analogue circuits. A popular waveform coding technique is pulse code modulation (PCM), which is widely used in telephony today.

Vocoders use an entirely different approach to speech coding, known as parameter-coding, or analysis-synthesis coding where no attempt is made at reproducing the exact speech waveform at the receiver, only a signal perceptually equivalent to it. These systems provide much lower data rates by using a functional model of the human speaking mechanism at the receiver. One of the most popular techniques for analysis-synthesis coding of speech is called Linear Predictive Coding (LPC). Some higher quality vocoders include RELP (Residual Excited Linear Prediction) and CELP (Code Excited Linear Prediction)

The problem at hand is to compress speech signals using wavelets to improve, an existing (lossless) PCM compression techniques. The speech signals that need to be compressed are narrowband signals with frequencies ranging from 0 to 4 kHz. The sampling frequency should be at 8 kHz. Another constraint is not to exceed an overall bit rate of 16 kbps (compression ratio 4:1). The performance of the developed algorithm is compared against the following compression schemes Linear Productive Coding (LPC) which reduces the transmitted data by factor of more than twelve, and Global System Mobile (GSM) which reduces the transmitted data by factor of five.

1.1 Motivation for Studying this Problem

Speech is an acoustic waveform that conveys information from a speaker to a listener. When two parties are at a distance from each other there must be a medium to transmit the speech signals. There are two types of transmission: analog transmission and digital transmission. Uncompressed digital speech consumes a large amount of storage space and transmission bandwidth. Compressed digital transmission of speech is more versatile than analog, providing the opportunity of achieving lower costs, consistent quality and security. Digital speech coding or speech compression is concerned with obtaining compact digital representations of voice signals. The objective in speech coding is to represent speech with a small number of bits while maintaining its perceptual quality. Perceptually irrelevant information in the speech signal makes it possible to encode speech at low bit-rates. The capability of speech compression has been central to the technologies of robust long-distance communications and high-quality speech storage. Compression continues to be a key technology in communications in spite of the promise of optical transmission media of relatively unlimited bandwidth. This is because of our continued and, in fact, increasing needs to use band-limited media such as radio and satellite links. Furthermore, storage and archival of large volumes of spoken information makes speech compression essential even in the context of significant increases in the capacity of storage.

Since the bandwidth of a signal is a function of its bit-rate, low bit-rate speech technology is a key factor in meeting the increasing demand for new digital wireless communication services. Impressive progress has been made during recent years in coding speech with high quality at low bit-rates and at low cost. The rapid advancement in the efficiency of digital signal processors and digital signal processing techniques has stimulated the development of speech coding algorithms. These trends entail a continued interest in speech compression technology as they provide a viable means to realize reduced operating costs in voice communication systems.

1.2 Objectives:

The main objective of this research is to design and implement a low-bit rate speech compressor using wavelet decomposition. The major issues concerning the design of this Wavelet based speech coder are:

- To assign suitable wavelet filter for speech compression.
- To apply global and by level dependent threshold technique.
- To determine the effect of decomposition levels on both global and local threshold technique.
- To apply scalar quantization and Huffman coding in order to achieve high compression and good quality.
- To compare the obtained result with Linear Productive Coding (LPC) and Global System Mobile (GSM).

1.3 Organization of the Thesis

The thesis has been organized as follows:

Chapter 1 gives a general background and overview of speech compression.

Chapter 2 covers the literature review, including highlights of some speech coding techniques and the compression algorithms currently in use. These algorithms include both lossless and lossy methods. Furthermore, the wavelet theory is introduced.

Chapter 3 consists of the methodology used in designing the proposed algorithm and explained how wavelet transform is applied to speech signals.

Chapter 4 analyzes the compression results, with the inclusion of local and global threshold techniques in the speech compression. Finally the chapter ends with a comparison between the proposed algorithm and some past techniques.

In **Chapter 5** an effective conclusion of the main research points, in addition to some future recommendations and suggestions have been presented.

CHAPTER II

LITERATURE REVIEW

Speech compression can be divided into two different techniques. The first is through the frequency domain and the other is through the time domain. The types of speech compression algorithm used depend on the functionality required and affects the quality of the output. Signal compression is done through the removal of redundancies. In the case of speech compression, further compression can be obtained by removing irrelevancy information in the speech signal. The goal in most compression technique is to reduce the transmitted data rate and the storage space required (Chen, et al. 2001).

In this chapter, the speech production is explained and a review of the different speech coding techniques will be proposed and analyzed,. Some common speech compression algorithms (widely used in today's technology) will be presented and the theory of wavelet transform will be discussed in detail, in addition to comparing wavelet with the Fourier transform.

2.1 Human Speech Production

Speech coding algorithms can be made more efficient by removing the irrelevant information from speech signals. In order to design a speech-coding algorithm, it is thus necessary to know about the production of human speech, its properties and human perception of the speech signals, so that the redundancies and the irrelevant parts of these signals can be identified.

A speech signal is produced in three stages: first of all, air flows outward from the lungs, the air flow is modified at the larynx, and finally, further constriction of the airflow occurs by varying the shape of the vocal tract (Akmajian, et al. 1984). Each sound has its own positioning of the vocal tract articulators (vocal cords, tongue, lips, teeth, velum and jaw). In the case of vowels, the airflow is unrestricted through the vocal tract while in the case of consonants the airflow is restricted at some points. Sounds can be classified further as voiced or unvoiced.

The vocal tract is modelled as a time varying filter. It amplifies certain sound frequencies and attenuates other frequencies. The sound is produced when a sound source excites the vocal tract filter. If the source is periodic, it produces voiced speech; and if the source is non periodic or noisy, it produces unvoiced speech (Agbinya 1996). The sound source occurs in the larynx and the base of the vocal tract, where the air flow can be interrupted by the vocal folds.

The periodic opening and, closing of the vocal cord, results in a periodic sound source or excitation. In the case of unvoiced speech the air is forced through a narrow constriction at some points in the vocal tract, and creates turbulence. The excitation is noise-like and typically has low energy.

2.2 Overview of Speech Coding

The main speech coding techniques are briefly discussed in this section. In order to simplify the description of speech coders they are often broadly divided into three classes: waveform coders, source coders and hybrid coders. Typically waveform coders are used at high bit rates (low compression ratio), and give very good quality speech. Source coders operate at very low bit rates (high compression ratio), but tend to produce synthetic speech. Hybrid coders use techniques from both source and waveform coding, and give good quality speech at intermediate bit rates. Figure 2.1 shows the typical behaviour of the speech quality versus bit-rate curve for the three main classes of speech coders.

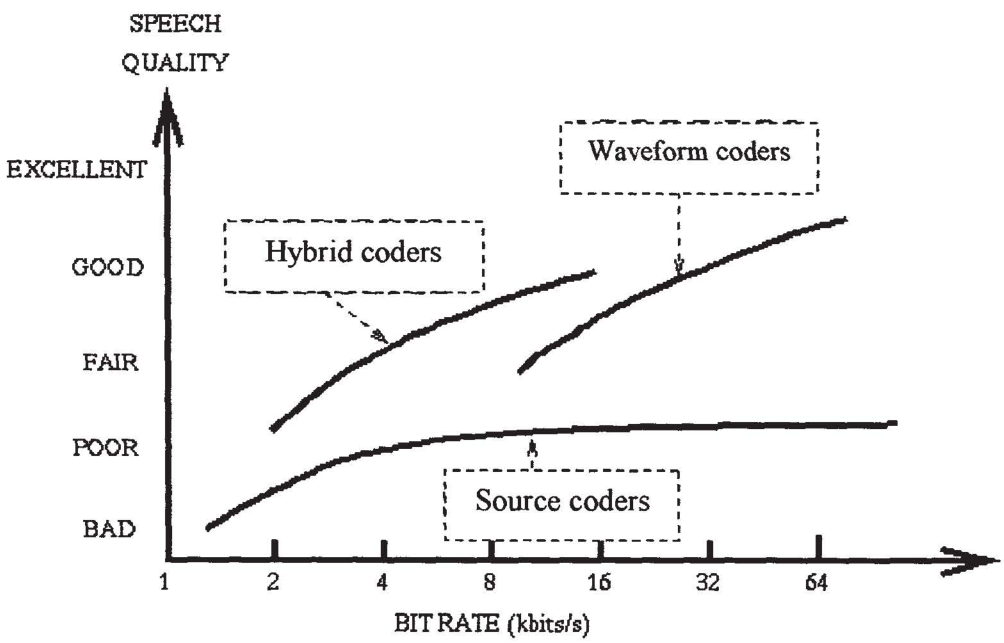

Figure 2.1 Quality comparison of speech coding schemes (Kondoz 1995)

2.2.1 Waveform Coders

Waveform coders attempt to reproduce the input signal waveform. They are generally designed to be signal independent so they can be used to code a wide variety of signals. Generally they are low complexity coders, which produce high quality speech at rates above about 16 kbps. Waveform coding can be carried out in either the time or the frequency domains (Spanias 1994).

2.2.1.1 Time Domain Coders

Time domain coders perform the coding process on the time samples of the signal data. The well known coding methods in the time domain are Pulse Code Modulation (PCM), Adaptive Pulse Code Modulation (APCM), Differential Pulse Code Modulation (DPCM), Adaptive Differential Pulse Code Modulation (ADPCM), Delta Modulation (DM), Adaptive Delta Modulation (ADM), and Adaptive Predictive Coding (Gersho.1994).

2.2.1.2 Frequency Domain Coders

Frequency domain waveform coders split the signal into a number of separate frequency components and encode them separately. The number of bits used to code each frequency component can be varied dynamically. Frequency domain coders are divided into two groups sub-band coders and transform coders.