

DEVELOPMENT OF A REAL-TIME SPEAKER RECOGNITION SYSTEM USING
TMS320C31

PRAKASH A/L SOWNDAPPAN

UNIVERSITI TEKNOLOGI MALAYSIA

DEVELOPMENT OF A REAL-TIME SPEAKER RECOGNITION SYSTEM USING
TMS320C31

PRAKASH A/L SOWNDAPPAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Engineering (Electrical)

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

MARCH 2006

To my beloved mother and father.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Ir. Prof. Dr. Sheikh Hussain Shaikh Salleh for his guidance and motivation. Without his continued support and interest, this thesis would not been the same as presented here.

I am grateful to the staff of Pusat Pengajian Siswazah and the Faculty of Electrical Engineering for their assistance during my study. A special thanks for the lab assistants for their support and help.

My fellow postgraduate students should also be recognized for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members.

ABSTRACT

Speaker recognition systems based on Malay language have been developed in the personal computer environment. This thesis outlines a hardware implementation of a real-time speaker recognition using Malay language. Various speaker recognition classifiers have been investigated in term of feasibility in a stand-alone hardware platform. Computational and memory requirement are given consideration, along with processing optimizations. A speaker recognition board is implemented based on a TMS32C31 digital signal processor (DSP). The speaker recognition techniques used are the Linear Predictive Coding (LPC) Cepstral analysis for feature extraction, Vector Quantization (VQ) for feature compression and the Dynamic Time Warping (DTW) for speaker feature matching. This system is trained and tested using a population of ten users, with additional testing using ten impostors. The average entry success of a true user is 93.4%. The speaker recognition board is successfully tested as a speaker recognition door access system, with true access success rate of 88.7%. The speaker recognition system shows good performance, as well as being operational in real-time.

ABSTRAK

Pengecaman suara berdasarkan bahasa Melayu telah dibangunkan berasaskan komputer peribadi. Tesis ini mengkaji perkakasan pengecaman suara bahasa Melayu yang dapat beroperasi dalam masa nyata. Pelbagai teknik pengklasifikasi pengecaman suara dikaji dari segi kebolehan operasi dalam perkakasan yang dapat beroperasi dengan sendiri. Keperluan komputasi and ingatan dipertimbangkan, beserta dengan optimasi pemprosesan. Perkakasan berasaskan pemproses isyarat digital (DSP) TMS32C31 telah dibangunkan. Teknik pengemcaman suara yang digunakan ialah Linear Predictive Coding (LPC) Cepstral, Vector Quantization (VQ) dan Dynamic Time Warping (DTW). Sistem ini telah diajar and diverifikasi menggunakan sepuluh pengguna, beserta sepuluh lagi orang bukan pengguna sistem. Purata keberkesanan memperolehi kebenaran laluan oleh pengguna ialah 93.4%. Perkakasan yang dibangunkan berjaya diuji sebagai sistem pengecaman suara untuk laluan pintu, dengan keberkesanan memperolehi kebenaran laluan 88.7%. Perkakasan yang dibangunkan menunjukkan tahap operasi yang memuaskan, dan dapat beroperasi dalam masa nyata.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
1	INTRODUCTION	1
1.1	Overview of Speaker Recognition	1
1.2	Real-Time Speaker Recognition System	2
	1.2.1 Related Work	2
	1.2.2 Objective	2
	1.2.3 Scope of Work	3
1.3	Organization of the Thesis	4
2	FUNDAMENTALS OF SPEAKER RECOGNITION	5
2.1	Basic Model of Speaker Recognition	5
2.2	Sampling and Preprocessing	7
	2.2.1 Analogue-To-Digital (ADC) Conversion	7
	2.2.2 Preemphasis	8
	2.2.3 Speech Framing	9
	2.2.4 Edge Detection	10
	2.2.5 Windowing	12
2.3	Feature Extraction	13
	2.3.1 The LPC Model	14
	2.3.2 Autocorrelation Analysis	14
	2.3.3 LPC analysis	14
	2.3.4 Cepstral Analysis	15
2.4	Vector Quantization	16
	2.4.1 VQ Codebook Design	17
2.5	Feature matching and speaker classification	19

2.5.1	Dynamic Time Warping	19
2.5.2	Vector Quantization	21
2.5.3	Hidden Markov Model	22
2.5.3.1	Problem 1 – Probability Evaluation	23
2.5.3.2	Problem 2 – Optimal State Sequence	24
2.5.3.3	Problem 3 – Parameter Estimation	25
2.5.4	Artificial Neural Network	26
2.6	Summary of the Algorithms Chosen for the Speaker Recognition System	27
3	DEVELOPMENT OF ALGORITHM FOR REAL- TIME IMPLEMENTATION	30
3.1	Speech Sampling	30
3.2	Edge Detection	32
3.3	Autocorrelation Analysis	34
3.3.1	Offline Autocorrelation Computation	35
3.3.2	Offline Autocorrelation Computation at the End of Each Frame	36
3.3.3	Inline Autocorrelation Computation	37
3.3.4	Inline Autocorrelation Computation for Overlapping Speech Frames	39
3.3.5	Autocorrelation computation comparison	43
3.3.6	Summary	44
3.4	Computation and Memory Requirement for Various Classifiers	45
3.4.1	Vector Quantization	46
3.4.2	Dynamic Time Warping	46
3.4.3	VQ-DTW	47
3.4.4	Hidden Markov Models	47
3.4.5	Artificial Neural Networks	48
3.4.6	Speaker Recognition Classifier Feasibility	49

4	SYSTEM DESIGN	52
4.1	Hardware Design	52
4.1.1	Digital Signal Processor	53
4.1.2	Memory	54
4.1.3	Analog Interface	57
4.1.4	Real-Time Clock	58
4.1.5	Power and Heat Consideration	59
4.1.6	PCB Design	59
4.2	Software Design	59
4.2.1	Software Development Flow	60
4.2.2	Software Modules	62
4.3	Recognition Threshold	64
4.4	Combination Lock Numbers for Speaker Verification	64
4.4.1	Single Number Authentication	66
4.4.2	Combination Number Authentication	
	– Method 1	67
4.4.3	Combination Number Authentication	
	– Method 2	67
4.4.4	Varied Threshold Values	67
4.4.5	Combination Lock Observation	69
5	RESULTS AND DISCUSSION	71
5.1	Test Methodology	71
5.2	Test Results	73
5.2.1	True Speaker Access Result	73
5.2.2	Impostor Access Result	74
5.2.3	Training and Access Timing	75
5.3	Application as a Door Access System	76
5.3.1	System Overview	76
5.3.2	System Performance	77
5.4	Comparison with Commercial Product	77

5.5	Summary	78
6	CONCLUSION	80
6.1	Summary	80
6.2	Limitation of the Design	81
6.3	Suggestion for Further Work	81
	REFERENCES	83
	Appendices A - F	86 - 97

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Typical frame length and window shift for different sampling frequency	10
2.2	Summary of common windowing function	12
3.1	Speech processing parameters	36
3.2	Processing and storage estimation for methods of autocorrelation computation	43
3.3	Simulation result on the processing and storage requirement for the different methods of autocorrelation computation	44
3.4	Parameter storage area for HMM	48
3.5	Storage and processing time for different classifiers	50
4.1	Comparison of various memory devices	55
4.2	Performance comparison of different optimization level	61
4.3	Probability of successful verification of 3-number combination lock	65
4.4	Profile of speakers	65
4.5	True speaker verification result	66
4.6	Impostor verification result	66
4.7	False rejection and false acceptance as threshold value carried over $0.5T - 2T$	68
5.1	Profiles of 10 enrolled users	72
5.2	Profiles of 10 impostors	73

5.3	True speaker identification and verification success count	74
5.4	True speaker entry attempts success count	74
5.5	Impostor recognition result	75
5.6	Entry attempts by impostor	75

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Block diagram of the speech recognition system	3
2.1	Basic model of a speaker identification system	6
2.2	Basic model of a speaker verification system	6
2.3	Blocking of speech frames into overlapping frames	9
2.4	Edge detection of sampled speech signal	10
2.5	Flowchart of the edge detection algorithm	11
2.6	The hamming window plot	13
2.7	Block diagram of the basic VQ structure	17
2.8	VQ codebook organization	17
2.9	Flowchart of the binary split codebook generation algorithm	18
2.10	Example of time alignment between two speech utterances	20
2.11	Example of DP recursion with constrain	21
2.12	A simple 5 state HMM	22
2.13	Simple computation element of an artificial neuron	26
2.14	A three-layer ANN	27
2.15	Processing method for the speaker recognition implementation	29
3.1	Processor utilization of DSP idling during sampling	32
3.2	Processor utilization of DSP executing code during sampling	32
3.3	Sample of speech utterance during verification	33

3.4	Flowchart of the edge detection algorithm	34
3.5	Autocorrelation computation for M=3 and N=15	35
3.6	Inline autocorrelation computation for M=3 and N=15	38
3.7	Inline autocorrelation computation for overlapping frames, with no windowing function	41
3.8	Inline autocorrelation computation for overlapping frames, with hamming windowing function applied	42
4.1	Block diagram of the hardware	52
4.2	Memory map of the speaker recognition board	56
4.3	The designed speaker recognition board	59
4.4	Software development flow for the speaker recognition system	60
4.5	Block diagram of software modules developed for the speaker recognition system	63
4.6	False rejection and acceptance for a one number authentication	68
4.7	False rejection and acceptance for 3- number authentication (method 1)	69
4.8	False rejection and acceptance for 3- number authentication (method 1)	69
5.1	The door entry interface panel	77

LIST OF SYMBOLS AND ABBREVIATIONS

ADC	-	Analog-to-digital conversion
AIC	-	Analog interface circuits
ALU	-	Arithmetic logic unit
ANN	-	Artificial neural networks
ASIC	-	Application Specific Integrated Circuit
CCS	-	Code Composer Studio
COFF	-	Common object file format
DAC	-	Digital-to-analog converter
DMA	-	Direct memory access
DNL	-	Differential nonlinearity
DP	-	Dynamic programming
DSK	-	DSP Starter Kit
DSP	-	Digital signal processor
DTW	-	Dynamic Time Warping
EER	-	Equal error rate
EPROM	-	Electrically programmable read-only memory
EEPROM	-	Electrically erasable programmable read-only memory
FPGA	-	Field Programmable Generic Array
GMM	-	Gaussian Mixture Model
HMM	-	Hidden Markov Models
IC	-	Integrated circuit
I/O	-	Input/output
IER	-	Identification error rate
K	-	Number of overlapping frames
KNN	-	K-Nearest Neighbor

L	-	Frame count
LCD	-	Liquid crystal display
LPC	-	Linear Predictive Coding
M	-	Window shift
MFCC	-	Mel-Frequency Cepstrum Computation
MFLOPS	-	Million floating-point operations per second
MIPS	-	Million instructions per second
N	-	Frame length
NV	-	Non-volatile
PC	-	Personal computer
PCB	-	Printed circuit board
QFP	-	Quad flatpack
RAM	-	Random access memory
RCC	-	Real Cepstral Coefficients
ROM	-	Read-only memory
RTC	-	Real-time clock
SMT	-	Surface mount technology
SNR	-	Signal to noise ration
SQRR	-	Signal to quantization noise ratio
VQ	-	Vector quantization
a	-	Degree of preemphasis

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Schematic diagram	86
B	Composite drawing of PCB Layout	92
C	Sampling loop and interrupt service routine	93
D	Autocorrelation analysis source code	94
E	LPC cepstrum analysis source code	95
F	DTW source code	97

CHAPTER 1

INTRODUCTION

In this chapter, an overview into the area of speaker recognition is given. Related work on real-time speaker recognition system, along with the objective and scope of work are given, and ended by a short description of the chapters of this thesis.

1.1 Overview of Speaker Recognition

Personal identity identification is an important requirement to controlling access to protected resources. Personal identity is usually claimed by using personal possessions like a key or badge, or knowledge of certain information like a password or combination numbers (Naik, 1990). However, these can be lost, stolen or counterfeited, thereby posing a threat to security. Biometric identification by using certain features of a person is a more secured solution for security identification. Fingerprint, hand geometry, retinal scanning, speech, and handwriting are examples of biometric features.

Speaker recognition is a process of identifying a person on the basis of speech processing. Speaker recognition can be more precisely described as the use of a machine to recognize a person from a spoken phrase (Campbell, 1997). Advances in speech processing technology and digital signal processors have made possible the design of high-performance and practical speaker recognition systems. Speaker recognition can be used for secured applications like phone banking, voice mail, door access, and access of computer networks. Speaker recognition has also found usefulness in forensic application (Champod and Weuwly, 2000).

Speaker recognition has been shown to yield near perfect recognition results. Speaker identification using 630 speakers from the TIMIT database has given a result of less than 0.5% error rate (Reynolds, 1995). IER (identification error rate) of less than 0.75% have been reported during speaker identification and EER (equal error rate) of less than 0.13% during speaker verification experiments (Wildermoth and Pawilal, 2003). These results were obtained using the TIMIT, YOHO and ANDOSL databases.

1.2 Real-Time Speaker Recognition System

For a speaker recognition system to be used in practice, the response time, or the time taken to train and to recognize the speakers must be minimized to an acceptable level. The size of the speaker database grows when the number of speakers in a system is increased. This poses two problems in terms of memory requirement for speech database storage, and processing time required by the system.

1.2.1 Related Work

Work by Karpov (2003) on real-time speaker identification system has been concentrated in the area of algorithm optimization, without looking into the hardware design aspect. Sara (1998) concentrated on optimized speech processing in the DSP56001 hardware platform, especially in the application of noise reduction and speech enhancement. Kwek (2000) worked on a hardware based speech recognition system much similar to the hardware developed in this master's work. Both work by Sara (1998) and Kwek (2000) are hardware based but are not concentrated in the area of speaker recognition.

1.2.2 Objective

The objective of the work is to investigate the feasibility and to implement a hardware based real-time speaker recognition system.

1.2.3 Scope of Work

A significant amount of experiments and study have been outlined in this thesis in these two areas: processing and storage requirement for real-time speaker recognition system. Description of a TMS320C31 based speaker recognition system is given. This system is tested in a real application as a speaker recognition door access system. An inline computation method of autocorrelation computation that reduces recognition time is benchmarked in terms of processing and memory requirement. The feasibility study of using various feature matching techniques in a real-time system is also outlined in terms of processing and memory requirement. The effect of combination lock number in speaker verification is investigated.

Figure 1.1 shows the block diagram of the final implementation of the speech recognition hardware implementation. A TMS320C31 digital signal processor running at 50MHz is used to execute the speaker recognition algorithm. The LPC Cepstral technique is used for feature extraction of speech signal. The classifier used is the combination of VQ-DTW feature compression and matching techniques.

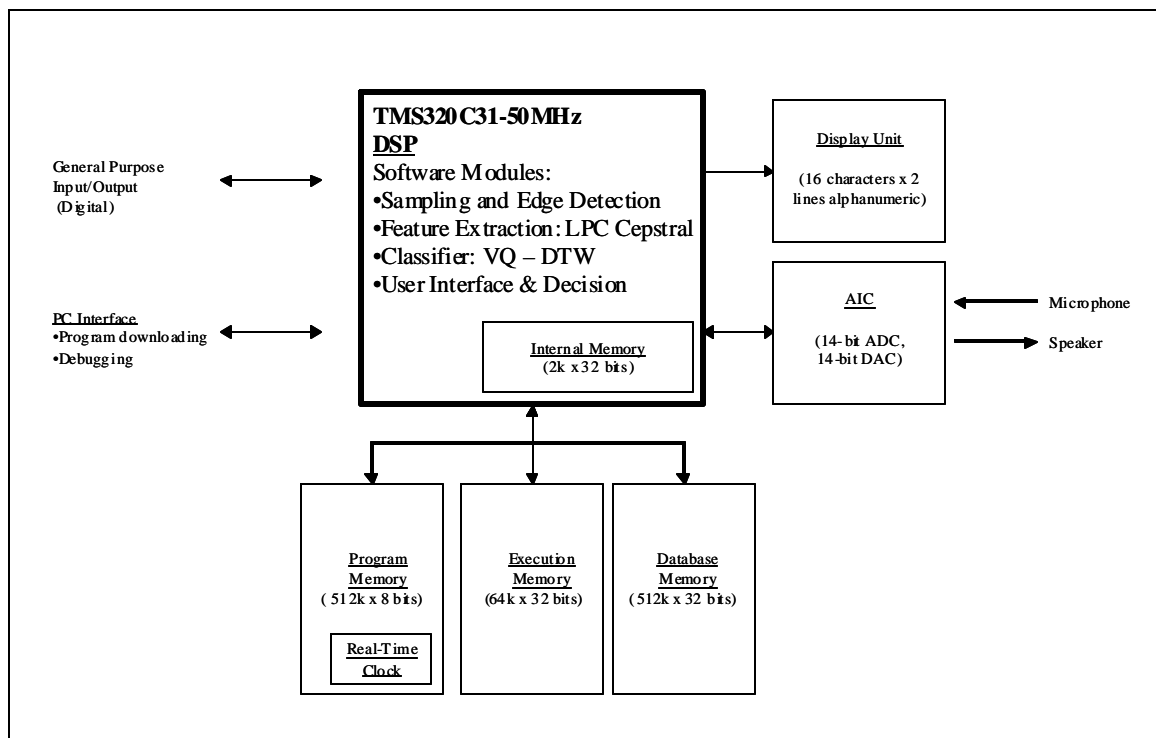


Figure 1.1: Block diagram of the speech recognition system

1.3 Organization of the Thesis

In Chapter 2, a brief explanation of various algorithms and approaches used in speaker recognition systems are given. Speaker recognition starts with speech acquisition, followed by speech feature extraction, and the modeled for speaker discrimination. Feature comparison and modeling techniques like Dynamic Time Warping, Vector Quantization, Hidden Markov Models, and Artificial Neural Networks are explained.

Chapter 3 examines the algorithms and various optimizations used in the implementation of a real-time speaker recognition system. A feasibility study on speaker recognition modeling algorithms is outlined here.

The hardware and software implementation of the speaker recognition system are outlined in Chapter 4.

The results of experiments and discussion of the developed speaker recognition system are given in Chapter 5.

The final chapter, Chapter 6 summarizes the research findings, the contributions and the limitations of the design. This chapter also outlines the direction for future work.

The limitation of the design is given in Section 6.2. Bandpass filtering and noise reduction will make the system more robust to noise. The implementation of word spotting will make the system more robust and eliminates the need for pauses between digits of the combination phrase.

REFERENCES

- Atal, B. S. (1974). Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification. *J. Acoust. Am.* 55(6): 1304 - 1312.
- Battle, E. Fonollosa, J. (1996). CPU and Memory Requirements for Real-Time Speech Recognition Systems Using the TMS320C3x/C4x. *Proc. The First European DSP Education and Research Conference Paris (France)*.
- Bennani, Y. and Fogelman, F. (1990). A connectionist approach for automatic speaker recognition. *Proc. ICASSP 90*. April. Albuquerque, Mexico. 265 - 268.
- Campbell, J.P. (1997). Speaker recognition: a tutorial. *Proc. IEEE*. 85(9): 1437 - 1462.
- Champod, C. and Meuwly, D. (2000). The inference of identity in forensic speaker recognition. *Speech Communication*. 31(2-3): 193 - 203.
- Karpov, Evgeny. (2003). *Real-time speaker identification*. University of Joensuu: Master Thesis.
- Kester, Walt. (2004). *Analog-Digital Conversion*, U.S.: Analog Devices.
- Kinnunen, T. and Fränti, P. (2001). Speaker Discriminative Weighting Method for VQ-Based Speaker Identification. *Proc. 3rd International Conference on audio- and video-based biometric person authentication (AVBPA)*. June 6-8, Halmstad, Sweden, 150 - 156.
- Kinnunen, T. and Kärkkäinen, I. (2002). Class-Discriminative Weighted Distortion Measure for VQ-Based Speaker Identification. *Proc. Joint IAPR International Workshop on Statistical Pattern Recognition (S+SPR 2002)*. August 6-9, Windsor, Canada, 681-688.
- Kwek, Ser Wee. (2000). *Real Time Implementation of Speech Recognition System Using TI Floating-Point Processor TMS320C31*. Universiti Teknologi Malaysia: Final Year Project Thesis.
- Markel, J. D. and Gray, A. H. (1976). *Linear Prediction of Speech*. Berlin, Germany.: Springer-Verlag.

- Naik, J.M. (1990). Speaker Verification: A Tutorial. *IEEE Communication Magazine*. January, 42-47.
- Oglesby, J. Mason, J. S. (1990). Radial basis function network for speaker recognition. *Proc. ICASSP 90*. April, Albuquerque, Mexico, 261-264.
- Oglesby, J. Mason, J. S. (1991). Radial basis function network for speaker identification. *Proc. ICASSP 91*. May, Toronto, Canada. 393-396.
- Parsons, Thomas W. (1987). *Voice and Speech Processing*. McGraw-Hill.
- Rabiner, Lawrence R. and Juang, B, H. (1986). An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, 3(1): 4-16.
- Rabiner, Lawrence R. and Juang, B, H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, N.J.: Prentice-Hall.
- Reynolds, D. (1995). Large population speaker identification using clean and telephone speech. *IEEE Signal Processing Letters*. 2: 46-48.
- Saito, Shuzo and Nakata, Kazuo. (1985). *Fundamentals of Speech Signal Processing*. U.S.: Academic Press.
- Sara, Grassi (1998). *Optimized implementation of speech processing algorithms*. University of Neuchatel: Ph.D. Thesis.
- Sensory, Inc. (2002). *RSC-4X Family Speech Recognition Microcontroller*. U.S.: 80-0220-A 2/02 Datasheet.
- Shao, Cheng Woo. (2000). *Speaker Recognition Using ANN*. Delft University of Technology: Master Thesis.
- Soong F. K., Rosenberg, A. E. and Rabiner, L. R. (1985). A Vector Quantization Approach to Speaker Recognition. *Conference Record 1985 IEEE International Conference on Acoustics, Speech, and Signal Processing*. March. 387-389.
- Texas Instruments. (2004). *SM320C6711-EP, SM320C6711B-EP, SM320C6711C-EP, SM320C6711D-EP Floating-Point Digital Signal Processors*. U.S.: SGUS054 Datasheet.
- Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Trans. Information Theory*, IT-13. April. 260-269.

Wildermoth, B. and Paliwal, K.K. (2003). GMM based speaker recognition on readily available databases. Proc. *Microelectronic Engineering Research Conference*. November. Brisbane, Australia.