UNIVERSITI PUTRA MALAYSIA

DEVELOPMENT OF AN ISOLATED DIGIT SPEECH RECOGNITION BASED ON MULTILAYER PERCEPTRON MODEL

UMMU SALMAH BINTI MOHAMAD HUSSIN

FSKTM 2004 7

DEVELOPMENT OF AN ISOLATED DIGIT SPEECH RECOGNITION BASED ON MULTILAYER PERCEPTRON MODEL

By

UMMU SALMAH BINTI MOHAMAD HUSSIN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

August 2004

DEVELOPMENT OF AN ISOLATED DIGIT SPEECH RECOGNITION BASED ON MULTILAYER PERCEPTRON MODEL

UMMU SALMAH BINTI MOHAMAD HUSSIN

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2004

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Doctor of Philosophy.

DEVELOPMENT OF AN ISOLATED DIGIT SPEECH RECOGNITION BASED ON MULTILAYER PERCEPTRON MODEL

By

UMMU SALMAH BINTI MOHAMAD HUSSIN

August 2004

Chairman : Associate Professor Ramlan Mahmod, Ph.D.

Faculty : Computer Science and Information Technology

The automatic speech recognition (ASR) field has become one of the leading speech technology areas nowadays. The research in ASR has always been emphasizing on developing man-machine communication and promising in ease of use over the traditional keyboard and mouse. The speech recognition task is simple to be identified by human, but a very complex process for the machine to understand. Various methods have been introduced to develop an efficient ASR system. A Neural Network (NN) approach is one of the famous methods and widely used in this field. A Multilayer perceptron (MLP) is a popular NN model used in ASR field. In this study, a MLP with back propagation learning algorithm is implemented to perform the isolated digit speech recognition task for Malay language. However, one of the current problems faced by MLP and most NN models in ASR field is the long learning time. Besides that, the requirement to produce high recognition rate for isolated digit speech recognition system performed by MLP is also not trivial because it has been widely used in many applications. Thus, this study focuses on improving the learning time and recognition rate of the MLP neural network for

Malay isolated digit speech recognition system. This current study proposes three new methods to fulfill the objective above. The improvement is made in preprocessing and recognition phase. In preprocessing phase, a new endpoint detection method is proposed and it is known as variance method. This method is introduced to overcome the disadvantages of the conventional method. The obstacles in the conventional method are unstable and difficult to set the threshold during the silence detection. Hence, poor recognition rate is produced. Another contribution in the preprocessing phase is in normalization phase. Three normalization methods are introduced to normalize the speech data before propagating to NN. The proposed methods consist of exponent, hybrid I and hybrid II. These methods are compared with 4 widely used conventional normalization methods. These include range I, range II, simple and variance method. The conventional methods have two limitations. The first is that some of the methods are very slow in learning phase but produce good recognition rate such as variance and range I methods. The second is that few of them are very fast in learning phase but produce low recognition rate such as simple and range II methods. Therefore, the new normalization methods are proposed to accelerate learning time and to produce high recognition rate. In recognition phase, a simple novel approach is introduced to increase the recognition rate. An adaptive sigmoid function is implemented to achieve this objective**.** A typical or fixed sigmoid function method is used in learning phase. In the recognition phase, an adaptive sigmoid function is employed. In this sense, the slope of the activation function is adjusted to gain highest recognition rate. This study emphasizes on 10 Malay words that comprise of “sifar” to “sembilan” (“0” to “9”). All utterances were recorded through single male speaker and each utterance was repeated 100 times. Thus the data set consist of 1000

utterances of Malay words. Four hundred data sets were split to utilize in the learning phase and the remaining 600 data for recognition phase. The TI46 standard data set was used to evaluate the performance of the all proposed method and 10 English words, consisting of "***zero***" to "***nine***" ("0" to "9") are utilized throughout this study. Eight male and female speakers uttered each word 8 times. Hence, the total data set is 1600 for both speakers. The data set based on male and female speaker is trained separately. In this sense, four hundred male data sets were experimented during learning phase; meanwhile 400 data sets are kept as test data. The same approach is utilized in learning and recognition phase for female data sets. The Linear Predictive Coding (LPC) is implemented as a feature extraction method to represent the speech data. The experimental results show that the proposed endpoint detection (variance method) produced promising results in term of learning time and recognition rate. Meanwhile, the proposed normalization method has shown excellent results over all experiments. The adaptive sigmoid function also successfully increased the recognition rate in the most of the experiments. Finally, from the overall experiments, it can be concluded that the highest recognition rate for Malay data set is 99.83% with 82s convergence time. Meanwhile, for TI46 data set (female and male data set), the yielded convergence time is 55s and 111s with the recognition rate of 96.75% and 94.75% respectively.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan ijazah Doktor Falsafah.

PEMBANGUNAN PENGECAMAN PERTUTURAN DIGIT TERPENCIL BERASASKAN MODEL MULTI ARAS *PERCEPTRON*

Oleh

UMMU SALMAH BINTI MOHAMAD HUSSIN

Ogos 2004

Pengerusi: Profesor Madya Ramlan Mahmod, Ph.D.

Fakulti : Sains Komputer dan Teknologi Maklumat

Bidang pengecaman pertuturan automatik telah menjadi salah satu bahagian teknologi pertuturan yang utama masa kini. Kajian terhadap pengecaman pertuturan sering menekankan pembangunan komunikasi manusia mesin dan menjanjikan pengendalian penggunaannya yang mudah berbanding dengan pendekatan tradisional papan kekunci dan tetikus. Pengecaman pertuturan mudah untuk dikenal oleh manusia, tetapi amat kompleks untuk difahami oleh mesin. Pelbagai kaedah telah diperkenalkan untuk membangun sistem pengecaman pertuturan automatik yang efisien. Rangkaian neural merupakan salah satu pendekatan terkenal yang sering digunakan dengan meluas di dalam bidang ini. Perseptron multi aras merupakan model rangkaian neural yang popular dalam bidang pengecaman pertuturan. Dalam kajian ini, *perseptron multi aras* dengan kaedah pembelajaran rambatan balik telah diimplementasi untuk melaksanakan tugas pengecaman pertuturan digit terpencil bagi bahasa Melayu. Walaubagaimanapun, salah satu masalah yang dihadapi oleh perseptron multi aras dan model rangkaian neural lain di dalam bidang pengecaman pertuturan adalah masa latihan yang terlalu lama. Di samping itu, ketepatan pengecaman pertuturan digit terpencil juga tidak boleh

diabaikan kerana ianya digunakan dengan meluas di dalam banyak aplikasi. Oleh itu, kajian ini memfokus pada pembaikan masa latihan dan ketegapan pengecaman bagi *perceptron* multi aras untuk memenuhi objektif di atas. Kajian semasa ini mencadangkan 3 kaedah baru untuk memenuhi objektif di atas. Pembaikan telah dilakukan di fasa pra pemprosesan dan pengecaman. Di fasa pra pemprosesan, kaedah pengesanan titik hujung yang baru dicadangkan dan dikenali sebagai kaedah varians. Kaedah ini diperkenalkan untuk mengatasi kelemahan pada kaedah konvensional. Salah satu kelemahan pada kaedah konvensional adalah kesukaran untuk menentukan nilai ambang ketika pengesanan bunyi senyap. Oleh yang demikian penghasilan ketepatan pengecaman adalah rendah. Sumbangan lain, di fasa pemprosesan adalah pada peringkat penormalan. Tiga kaedah penormalan diperkenalkan untuk menormalkan data pertuturan sebelum dihantar ke rangkaian neural. Kaedah dicadangkan terdiri daripada exponent, hybrid I dan hybrid II. Kaedah ini dibandingkan dengan 4 kaedah penormalan konvensional yang digunakan dengan meluas. Ini termasuk, range I, range II, simple dan kaedah varians. Kaedah konvensional mempunyai beberapa kekangan. Sebahagian daripada kaedah ini menghasilkan ketepatan pengecaman yang tinggi tetapi sangat lambat di dalam proses latihan. Sebaliknya, sebahagian daripadanya sangat laju di dalam masa latihan, tetapi menghasilkan ketepatan pengecaman yang rendah. Oleh itu, kaedah penormalan baru telah dicadangkan untuk melajukan masa latihan dan menghasilkan ketepatan pengecaman yang tegap. Di fasa pengecaman, pendekatan baru yang mudah diperkenalkan untuk meninggikan ketepatan pengecaman. Fungsi sigmoid menyesuai diimplementasi untuk mencapai objektif ini. Dalam pendekatan ini, fungsi sigmoid tetap digunakan pada proses latihan. Pada peringkat pengitlakan, fungsi sigmoid menyesuai digunakan. Dalam hal ini lereng fungsi aktivasi di ubahsuai untuk memperolehi ketepatan pengecaman yang tinggi. Kajian ini menumpukan pada 10 perkataan Melayu yang terdiri daripada “sifar” hingga “sembilan” (“0” hingga “9”). Kesemua sebutan ini telah dirakam dengan menggunakan suara jurucakap lelaki, iaitu setiap perkataan diulang sebanyak 100 kali. Oleh itu terdapat sebanyak 1000 perkataan sebutan Melayu. Empat ratus sebutan telah diasingkan untuk latihan manakala baki 600 digunakan untuk tujuan pengecaman. Set data piawai TI46 telah digunakan untuk mengukur prestasi kesemua kaedah yang dicadangkan dan 10 sebutan Inggeris yang terdiri daripada “***zero***’ hingga “***nine***” (“0” hingga “9”) telah digunakan sepanjang kajian. Lapan jurucakap lelaki dan jurucakap perempuan menyebut setiap perkataan sebanyak 8 kali. Oleh itu jumlah data adalah sebanyak 1600 bagi kedua-dua jurucakap. Set data bagi jurucakap lelaki dan perempuan di latih secara berasingan. Empat ratus data lelaki digunakan untuk latihan, manakala 400 lagi disimpan sebagai data ujian. Pendekatan yang sama digunakan untuk latihan dan pengujian bagi set data perempuan. Pemalar lelurusan telah diaplikasikan sebagai kaedah penyarian sifat bagi mewakili data pertuturan. Keputusan ujikaji menunjukkan titik hujung yang dicadangkan (kaedah varians) telah menghasilkan keputusan yang menyakinkan dari sudut masa dan ketepatan pengecaman. Manakala, kaedah penormalan yang dicadangkan telah memberikan keputusan yang cemerlang untuk semua ujikaji yang dijalankan. Fungsi sigmoid menyesuai berjaya juga telah berjaya meningkatkan ketepatan pengecaman bagi hampir keseluruhan ujikaji yang telah dilaksanakan. Akhir sekali, daripada keseluruhan eksperimen dapat disimpulkan bahawa oengecaman tertinggi untuk set data Melayu adalah 99.83% dengan masa penumpuan 82s. Sementara itu, bagi set data perempuan dan lelaki (TI46), masa penumpuan yang diperolehi adalah 55s dan 111s dengan ketepatan pengecaman masing-masing adalah 96.75% dan 94.75%.

ACKNOWLEDGEMENTS

Praise to **GOD** for giving me the opportunity to complete this study.

I would like to thank my supervisor Associate Professor Dr. Ramlan Mahmod for his guidance and assistance throughout my many years of studying in this program. Furthermore, his suggestion helped in making improvements in this work.

My deepest appreciation and gratitude to the committee members, Associate Professor Dr. Siti Mariyam Shamsuddin and Associate Professor Dr Abdul Rahman Ramli for help me with a lot of idea, motivation and encouragement during the study.

My deepest thanks also go to my family and friends for their supports and encouragement during the study.

I would thank to National Science Fellowship for giving the scholarship while I was carrying out this research.

Ummu Salmah Mohamad Hussin

August 2004

I certify that an Examination Committee met on 19 August, 2004 to conduct the final examination of Ummu Salmah Mohamad Hussin on her Doctor of Philosophy thesis entitled "Development of an Isolated Digit Speech Recognition Based on Multilayer Perceptron Model" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

ALI MAMAT, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

ABDUL AZIM ABD GHANI, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Examiner)

MD. NASIR SULAIMAN, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Examiner)

TENGKU MOHD. TENGKU SEMBOK, Ph.D.
Professor
Faculty of Technology and Information Science
Universiti Kebangsaan Malaysia
(Independent Examiner)

GULAM RUSUL RAHMAT ALI, Ph.D.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfilment of the requirements for the degree of Doctor of Philosophy. The members of the Supervisory Committee are as follows:

RAMLAN MAHMOD, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

SITI MARIYAM SHAMSUDDIN, Ph.D.
Associate Professor
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
(Member)

ABDUL RAHMAN RAMLI, Ph.D.
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

AINI IDERIS, Ph.D.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date :

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

UMMU SALMAH MOHAMAD HUSSIN

Date:

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADALINE	-	Adaptive Linear Neuron
ART	-	Adaptive Resonant Theory
ASR	-	Automatic Speech Recognition
BP	-	Backpropagation
BSB	-	Brain-State-In-A-Box
CMU	-	Carnegie Mellon University
DARPA	-	Defense Advanced Research Projects Agency
DTW	-	Dynamic Time Warping
HMM	-	Hidden Markov Model
IBM	-	International Business Machines Corporation
JSRU	-	Joint Speech Research Unit
LDC	-	Linguistic Data Consortium
LPC	-	Linear Predictive Coding
MIT	-	Massachusetts Institute of Technology
MLP	-	Multilayer Perceptron
NEC	-	Nippon Electronic Corporation
NN	-	Neural Network
OS	-	Operating System
RNN	-	Recurrent Neural Network
SRI	-	Stanford Research Institute

LIST OF FIGURES

LIST OF TABLES