SYSTEM FOR CHARACTERISATION AND RECOGNITION OF ARABIC PHONEMES AMONG MALAYSIAN CHILDREN USING FEED-FORWARD NEURAL NETWORKS

NURUL ASHIKIN BINTI ABDUL KADIR

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

JULY 2012

To my beloved
Embut Embong, Abdul Kadir A Rahman,
Nor Amin Rohimi
and
Rubita Sudirman (Dr.)

ACKNOWLEDGEMENT

Thank you Allah for giving me chance and good health experienced these precious moments doing my master study.

My greatest appreciations and thank you to my supervisor, Dr Rubita binti Sudirman for guiding and supervising me to complete this Master study successfully. My sincere appreciation to Ministry of Science, Technology & Innovation, Malaysia (MOSTI) and Universiti Teknologi Malaysia (UTM) for funding my Master study. I am also indebted to children from Sekolah Agama UTM and Umi Kalsom's Tuition Centre for being the research samples.

My sincere gratitude goes to the wonderful person in my life, Nor Amin Rohimi for being my helping hands and providing lots of assistance along this study completion. Also to Dr. Norlaili Mat Safri, Mr. Tan Chin Luh, Mrs. Husnaini Azmy, Mr. Afzan Othman, Dr Puspa Inayat Khalid who provide assistance along this study completion. Special thanks to Ms. Nor Hekmat Yahaya and Ms. Carolyn Foot for proofreading my abstract. Thanks for the valuable tips indeed. Also technicians from Medical Electronics Lab, Ms. Wan Nor Afiza whom always bothered by me, thank you so much.

My mom, dad, fiancé, sister, brothers and aunt who always support me along my way, thank you very much for their prayers, kind, love, care and patience.

ABSTRACT

There has been limited study and research in Arabic phoneme among Malaysians, hence making references to the work and research difficult. Although there have been significant acoustic and phonetic studies on languages such as English, French and Mandarin, to date there are no guidelines or significant findings on Malay language. This study discusses the correct use of Arabic phonemes pronunciation in Malay accent in the simplest form. The International Phonetic Alphabet (IPA) of Arabic chart is used as reference of every recorded speech sample using Malaysian children for sound localisation of every phoneme. Results from *Maahad Tahfiz* School teachers were used to identify the most suitable samples among the recorded samples. The samples were analysed to determine the origin of each phoneme data by measuring its formant frequencies. Consonants of Standard Arabic (SA) phonemes were studied and an appropriate place of articulation of every phoneme was measured through its formant. Seven out of 25 consonants of SA phonemes of the children's samples did not give the appropriate formants value were identified. The formant frequency values obtained were used as reference for the database for the proposed recognition system which was developed using Matlab software. All selected samples were randomly divided into 10 disjoint sets of equal size for validation, namely 10-fold cross validation, to estimate the performance of a predictive model. The mean square error (MSE) observed was 0.03, with the speech recognition using a developed neural network (NN) system. The results indicated that the highest training recognition rate obtained for multi-layer and cascade-layer network were 98.8 % and 95.2 % respectively, while the highest testing recognition rate achieved was 92.9 % for both networks and the MSE is 0.04.

ABSTRAK

Kajian mengenai penggunaan fonem bahasa Arab di kalangan rakyat Malaysia adalah sangat terhad dan ini menyukarkan rujukan dan kajian. Walaupun kajian akustik dan fonetik adalah sangat pesat untuk bahasa asing seperti Bahasa Inggeris, Bahasa Perancis dan Bahasa Mandarin namun hingga kini tiada panduan khusus didapati atau temuan yang penting dalam bahasa Melayu. Kajian ini dilaksanakan untuk mengkaji kesesuaian penyebutan fonem Arab dalam dialek Bahasa Melayu. Rujukan adalah bersandarkan carta Fonetik Huruf Antarabangsa (*IPA*) Bahasa Arab bagi memastikan ketepatan fonem yang direkod terhadap penyebutan setiap fonem. Keputusan daripada guru-guru sekolah Maahad Tahfiz digunakan dalam mengenalpasti sampel yang paling sesuai dari sampel yang telah direkodkan. Sampel-sampel tersebut dianalisa untuk menentukan titik asal setiap fonem dengan mengukur frekuensi-frekuensi formannya. Konsonan-konsonan fonem bahasa Arab Standard (*SA*) dikaji dan titik artikulasi sesuai setiap fonem diukur melalui formannya. Tujuh daripada 25 konsonan fonem *SA* bagi sampel kanak-kanak tidak memberikan nilai forman yang bersesuaian telah dikenalpasti. Nilai-nilai frekuensi forman yang diperoleh digunakan sebagai rujukan untuk pangkalan data bagi sistem pengecaman yang dicadangkan untuk dibina menggunakan perisian Matlab. Kesemua sampel terpilih dibahagi secara rawak kepada 10 set tidak berturutan yang sama saiz untuk pengesahan, iaitu 10-lipatan pengesahan bersilang, untuk anggaran prestasi model cadangan. Ralat min kuasa dua (*MSE*) diperoleh adalah 0.03, dengan pengecaman pertuturan menggunakan sistem jaringan saraf tiruan (*NN*) yang direka. Keputusan menunjukkan kadar latihan pengecaman paling tinggi diperoleh untuk jaringan lapisan pelbagai dan lapisan lata masing-masing adalah 98.8 % dan 95.2 %, sementara kadar ujian pengecaman diperoleh adalah 92.9 % untuk kedua-dua jaringan dan *MSE* ialah 0.04.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

There is not much research done on Arabic speech recognition which makes fewer references for the study compared to English. Furthermore, the researches done were on Arabian, Pakistan, Jordanian, Palestinian and other Arabian countries. Therefore, the purpose of this study is to develop an Arabic phonemes database and later will be used to develop software to show sound localization of Arabic letters recitation. This chapter tells about the history of speech recognition technology in brief, together with the problem statement of the study. Besides that, this chapter also discusses about the objectives of the study and the significance of the study.

1.2 Overview of Speech Technology Research History

The review of speech technology shows that the development of speech technology are developing and a lot of remarkable results have been produced as well as new challenges in the research fields. Nowadays, the application of speech processing technology (or known as Natural Language Processing Technology, NLP) have led to a number of successful applications available in software which suitable for personal-computer applications and hand-phones. NLP used via hand-phone, called smart-phone are attracting user interest as the *intelligent software assistant* with user-interface makes the technology such as Apple's and Google's on debut recently in the telecommunications world. NLP which manipulates the voice explicitly to produce signals that control computers or software or communication devices are user interest and research awareness.

As early as 50's, a system for isolated digit recognition for a single speaker was researched by Davis, Biddulph and Balashek (1952). Later, Olson and Belar (1956) designed a system to recognise 10-syllable for a single speaker. While in 1959, a system of recognising four vowels and nine consonants was built by Denes (1959) using spectral information in extracting voice features. By applying the same method, Forgie and Forgie (1959) built a speaker independent 10 English vowels recogniser in isolated words.

Throughout the 1960s, vowel recogniser, phoneme recogniser and digit recogniser developed. These researches were among Japanese, e.g. Suzuki and Nakata (1961), Sakai and Doshita (1962) and Nagata, Kato and Chiba (1964).

Afterward in year 1970s, most research focused on implementing techniques to improve the recognition technology. Sakoe and Chiba (1978) used warping function for spoken word normalization in spoken word recognition. Digits and names are considered in the recognition system. Itakura (1975) used Linear Predictive Coding (LPC) extraction method in speech recognition. While Rabiner *et al.* (1979) applied clustering techniques in the speaker independent recognition for isolated words.

Speech recognition research field in 1980's mostly employed of statistical modeling methods such as Neural Networks (NNs) and Hidden Markov Models (HMMs). The suitability of NN for pattern recognition was discussed by Lipmann (1987). Lippmann (1989) overviewed of clustering technique by using Multi-Layer Feed-Forward (MLFF) NN. The networks used Back-Propagation (BP) training rule effectively. As described by Juang and Furui (2000), statistical methods were designed to allow machine to learn the structure regularities in the speech signal directly from the data for the purpose of automatic speech recognition (ASR).

Later in year 1990's, shows a lot of additional improvement from the results developed years previously. Lee (1990) developed a continuous speech recognition speaker independent by using HMMs methods in context-dependent modeling. Liu *et al*., (1992) implemented Dynamic Time-Warping (DTW) with NN. Large vocabulary of continuous speech recognition was built using HMM toolkit, HTK by Woodland *et al*., (1994). HTK is a portable software toolkit developed by Cambridge University Speech Group. It uses continuous density HMMs for building speech recognition systems. Matsuura, Miyazawa and Skinner (1994) developed a speaker-independent isolated-word recognition system using NN to identify the fundamental sequence of phonemes and a DTW technique to time-align phonemes with equivalent lexical sequences of phonemes.

During 2000's, a broader research done in the field of speech recognition. Sutat and Chularat (2000) had developed Thai text-dependent speaker identification by using Multi-Layer Perceptron (MLP) trained with Back-Propagation (BP) algorithm. The used of linear interpolation and time-normalization are applied to the digits 0 to 9 in Thai language. Hong, Salleh and Sha'ameri (2002) developed speech recognition software for the purpose of teaching Linear Predictive Coding (LPC) and LPC-derived cepstrum (LPCC) in education environment. It provided with DTW, HMM and Vector-Quantization (VQ) techniques. Ting (2002) developed a system which concerned on the improvement over the limitations and problems of traditional speech therapy in Malay language. The speech classification used is multi-layer NN. Al-Sayegh and AbedEl-Kader (2004) developed an Arabic phoneme recogniser which used MLP NN method. While Pisarn and Theeramunkong (2007) developed a Thai spelling speech recogniser with HMM used in the system training phase. The Thai phonetic characteristics, alphabet system and spelling methods have been analysed. Grewal and Kumar (2010) designed an isolated word recognition system for English Language. Al-azzawi and Daqrouq (2011) developed Arabic vowel recognition based on wavelet and LPC by using Feed-Forward Back-Propagation Neural Networks (FFBPNN).

In spite of all of the progressions in the speech recognition field, a number of excellent commercial products are available, such as products that recognise human voice. IBM introduced their speech recognition system, Dragon (Pallet, 2002). Research describing about these technologies were published by Baker (1975), Radin (1983) and Atkinson and McCreight (1987). Also available through Google search engine, known as *Google's Voice Recognition* and available via Apple's smart-phone, *IPhone*, named as *Siri*.

1.3 Problem Statement

In Malaysia, still lack of Arabic speech recognition using neural networks. Most of the research focused on Malay language instead of Arabic language as the non-native language in their study. At present, a lot of Arabic software has been developed. Unfortunately, less research focused on children speech sounds as the children speech is more dynamic and inconsistent if compared to adult's speech (Ting and Yunus, 2004). Furthermore, children have shorter vocal tracts and smaller vocal cords thus having higher formant frequencies than adults (Boë *et al.*, 2006). Most speech recognition systems target to adult users and had experience severe speech recognition system performance degradation when used by children users. Zue *et al*. (2000) found that children's word error rate was almost twice adult users. In Malaysia's primary education, children were taught Arabic language as their third language after Malay language and English. Therefore this study focused on the characterisation, identification and classification of the children's Arabic speech sounds. Moreover, sound production point will be visualised using Matlab graphical user interface (GUI) for education purposes especially for children.

1.4 Objectives of the Study

There are four objectives for this study. The first objective is to identify the characteristics of Standard Arabic (SA) consonants. The characteristics of every SA consonants can be characterised by implementing Fast-Fourier Transform (FFT) and finding the formant frequencies from the signals.

The second objective is to employ *k*-fold cross validation technique to estimate the performance of the NN system recognition as the small network tend to learn the most gross behaviour of training data and ignore subtleties. *k*-fold cross validation is a reliable method on choosing the best NN after elimination of data by using formant frequencies pattern.

The third objective is to develop a system that can recognise Arabic phonemes sound pronunciation by using neural networks for pattern recognition and classification.

The forth objective is to design learning software for education purposes. A learning system is built by using the appropriate Arabic phonemes database and NN systems through visualising the sound production point at graphical user interface (GUI).

1.5 Scope of the Study

The limitation of this study is using children's voice as the database. The subjects are limited to healthy children aged 7 to 11 years. The formant frequencies of Standard Arabic (SA) consonants are studied as every phoneme has different formants pattern and characteristics.

This study focuses on SA consonants, which consist of 25 consonants. There are five manners of articulation for consonants which are fricative, plosive, trill, lateral and nasal are investigated. Their unique characteristics are identified as well, based on their formant frequencies.

1.6 Significance of the Study

Every Arabic phoneme possess different positions of the sound production organs of speech which comprise the tongue, mouth roof, teeth and the vocal cord which form the *makhraj* points to produce its pronunciation (refer to **APPENDIX A** for the Arabic letters sound production point). Despite that, the characteristics of Arabic phonemes among Malaysians children are identified based on their manners and places of articulation. Furthermore, the Arabic phonemes recognition system among Malaysian children is developed using neural networks. By visualising the *makhraj* point and produce a sample sound of each letters through graphical user interface (GUI), it can help children to get better understanding in their learning process of pronouncing the Arabic letters.

1.7 Organization of Thesis

This thesis is divided into five chapters. Following this introduction chapter is Chapter 2, which presents some background information of the study and literature review of related research of speech characteristics, speech recognition and neural networks.

In Chapter 3, we described briefly about methodology used in this study. Overall system design and implementation are described.

Chapter 4 discusses the findings during experiments and simulation of neural networks. Formant frequencies for every 25 consonants of Standard Arabic are determined. Further discussions about the best network selection were explained in this chapter.

Chapter 5 concludes research findings and recommendations of some ideas for future work.

REFERENCES

Ahmad A. M. and Ismail S. Recurrent Neural Network with Backpropagation Through Time For Arabic Recognition. Germany: SCS Europe. 13-16 June 2004.

Ahmad, A. M., Ismail, S. and Samaon, D.F. Recurrent Neural Network with Backpropagation through Time for Speech Recognition. *IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004*. 2004. 1: 98-102.

Al-azzawi, K.Y. and Daqrouq, K. Feed Forward Back Propagation Neural Network Method for Arabic Vowel Recognition Based on Wavelet Linear Prediction Coding. *International Journal of Advances in Engineering and Technology*. 2011. 1(4): 62-72.

Ali, A. M. A., Van der Spiegel, J. and Mueller, P. Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants. *IEEE Transactions on Speech and Audio Processing*. 2001. 9(8): 833-841.

Al-Sayegh, S. W. and AbedEl-Kader, A. F. Arabic Phoneme Recognizer based on Neural Networks. *Proceedings of International Conference on Intelligent Knowledge Systems, IKS2004*. Aug 16-20, 2004. Troy, Turkey. 2004.

Al-Shargabi, B., Al-Rominah, W. and Olayah, F. A Comparative Study for Arabic Text Classification Algorithms based on Stop Words Elimination. *ISWSA'11*. April 18-20, 2011. Amman, Jordan: ACM. 2011. Article 11.

Atkinson, R. R. and McCreight, E. M. The Dragon Processor. *ACM Sigarch Computer Architecture News*. 1987. 15(5): 65-69.

Awadalla, M., Abou Chadi, F. E. Z. and Soliman, H. H. Development of an Arabic Speech Database. *Enabling Technologies for the New Knowledge Society: ITI 3rd International Conference on Information and Communications Technology, 2005*. Dec 5-6, 2005. Cairo: IEEE. 2005. 89-100.

Awais, M. M., Masud, S., Akhtar, J. and Shamail, S. Arabic Phoneme Identification Using Conventional and Concurrent Neural Networks in Non Native Speakers. Lecture Notes in Computer Science. 2007. 4681: 897-905.

Awais, M. M., WaqasAhmad, Masud, S. and Shamail, S. Continuous Arabic Speech Segmentation using FFT Spectrogram. *Innovations in Information Technology 2006.* Nov 19-21, 2006. Dubai: IEEE. 2006. 1-6.

Bahi, H. and Sellami, M. Neural Expert Model Applied to Phonemes Recognition. In: Perner P. and Imiya A. *Lecture Notes in Computer Science Volume 3587/2005*. Germany: Springer-Verlag Berlin Heidelberg. 507-515; 2005.

Baker, J. K. The Dragon System – An Overview. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 1975. ASSP-23: 24-29.

Beautemps, D., Badin, P. and Laboissière, R. Deriving Vocal-Tract Area Functions from Midsagittal Profiles and Formant Frequencies: A New Model for Vowels and Fricative Consonants based on Experimental Data. *Speech Communication*. 1995. 16(1): 27-47.

Boë, L-J., *et al*. Skull and Vocal Tract Growth from Newborn to Adult. *Proceedings of the 7th International of Seminar on Speech Production, ISSP2006*. Dec 13-15, 2006. Ubatuba, Brazil: CEFALA. 2006. 75-82.

Bouckaert, R. R. Choosing between Two Learning Algorithms based on Calibrated Tests. *Proceedings of the Twentieth International Conference on Machine Learning, ICML 2003*. August 21-24, 2003. Washington DC: AAAI Press. 2003. 51-58.

Callan, R. *The Essence of Neural Networks*. Maylands Avenue: Prentice Hall. 1999.

Chandra, B. and Varghese, P. P. Applications of Cascade Correlation Neural Networks for Cipher System Identification. *World Academy of Science, Engineering and Technology 26*. Dec 14-16, 2007. Bangkok, Thailand: WASET. 2007. 312-314.

Cutler, A., Dahan, D. and Donselaar, W. Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech April/June 1997*. 1997. 40(2): 141-201.

Daniel, R. P. and Marcos, F. Z. Speaker Recognition with a MLP Classifier and LPCC Codebook. *IEEE International Conference on Acoustics, Speech & Signal Processing 1999*. 1999. 2: 1005-1008.

Davis, K. H., Biddulph, R. and Balashek, S. Automatic Recognition of Spoken Digits. *Journal of the Acoustical Society of America*. 1952. 24(6): 637-642.

Debyeche, M., Houacine, A. and Haton, J. P. A Knowledge-Based Approach for Arabic Amphatic Consonant Identification Based on Speech Spectrogram Reading. *Proceedings of the Thirtieth Southeastern Symposium System Theory, SSST 1998*. March 8-10, 1998. Morgantown, West Virginia: IEEE. 1998. 325-328.

Denes, P. The Design and Operation of the Mechanical Speech Recognizer at University College London. *Journal of the British Institution of Radio Engineers*. 1959. 19(4): 219-229.

Dietterich, T. G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algortihms. *Neural Comput*. 1998. 10(7): 1895-1923.

Fackrell, J. W. A., Vereecken, H., Martens, J-P., and Coile, B. V. Multilingual Prosody Modelling using Cascades Regression Trees and Neural Networks. *Sixth European Conference on Speech Communication and Technology EUROSPEECH 1999*. Sept 5-9, 1999. Budapest, Hungary: EUROSPEECH. 1999. 1835-1838.

Fahlman, S. E. An Empirical Study of Learning Speed in Back-Propagation Networks. Technical Report # CMU-CS-88-162. Carnegie Mellon University, Pittsburgh. September 1988.

Fahlman, S. E. and Lebiere, C. The Cascade-Correlation Learning Architecture. Technical Report # CMU-CS-90-100. Carnegie Mellon University, Pittsburgh. August 1991.

Fant, G. *Acoustic Theory of Speech Production with Calculations Based on X-Ray Studies of Russian Articulations*. Netherlands: Mouton and Co, The Hague. 1960.

Forgie, J. W. and Forgie, C. D. Results Obtained from a Vowel Recognition Computer Program. *Journal of the Acoustical Society of America*. 1959. 31(11): 1480-1489.

Gordon, M., Barthmaier, P. and Sands, K. A Cross-Linguistic Acoustic Study of Voiceless Fricatives. *Journal of the International Phonetic Association*. 2002. 32(2): 141-174.

Grewal, S. S. and Kumar, D. Isolated Word Recognition System for English Language. *International Journal of Information Technology and Knowledge Management*. 2010. 2(2): 447-45.

Hassan, A. *Linguistik Am Untuk Guru Bahasa Malaysia*. 5th ed. Selangor, Malaysia: Fajar Bakti. 1984.

Heinzel, G., Rudiger, A. and Schilling, R. Spectrum and Spectral Density Estimation by the DFT Including A Comprehensive List of Window Functions and Some New Flat Top windows. Internal Report, Max-Planck-Institut für Gravitationsphysik, Hannover. Feb 15, 2002.

Hong, K. S., Salleh, S-H. and Sha'ameri, A. Z. A New Developed Speech Recognition Education Software in Teaching LPC and LPC-derived Cepstrum. *Proceeding of ICSP '02*. 2002. 2: 1701-1704.

Iqbal, H. R., Awais, M. M., Masud, S. and Shamail, S. On Vowels Segmentation and Identification Using Formant Transitions in Continuous Recitation of Quranic Arabic. In: Nguyen, N. T. and Katarzyniak, R. *New Challenges in Applied Intelligence Technologie Vol. 134/2008*. Chennai, India: Springer-Verlag Berlin Heidelberg. 155-162; 2008.

Itakura, F. Minimum Prediction Residual Applied to Speech Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1975. ASSP-23: 67-72.

Jackson, P. J. B. Acoustic Cues of Voiced and Voiceless Plosives for Determining Place of Articulation. *Proc. Workshop on Consistent and Reliable Acoustic Cues for sound analysis, CRAC 2001*. September, 2001. Aalborg, Denmark: Eurospeech. 2001. 19-22.

Jackson, P. J. B. *Characterisation of Plosive, Fricative and Aspiration Components in Speech Production*. Ph.D. Thesis. Universiti of Southampton, UK; 2000.

Jassem, Acoustical Description of Voiceless Fricatives in terms of Spectral Parameters. *Speech Analysis and Synthesis 1*. 1967. 1: 189-206.

Jesus, L. M. T. and Jackson, P. J. B. Frication and Voicing Classification. *Computational Processing of the Portuguese Language, Lecture Notes in Computer Science*. 2008. 5190/2008: 11-20.

Jesus, L. M. T. and Shadle, C. H. Acoustic Analysis of European Portuguese Uvular [χ,ʁ] and Voiceless Tapped Alveolar [ɾ̥] Fricatives. *Journal of the International Phonetic Association*. 2005. 35(1): 27-44.

Johnson, D. Modeling the Speech Signal. Version 2.28. 2009. http://cnx.org/content/m0049/2.28/

Johnson, K. *Acoustic & Auditory Phonetics*. 2nd ed. United Kingdom: Blackwell Publishing. 2003.

Juang, B. H. and Furui, S. Automatic Recognition and Understanding of Spoken Language – A First Step Toward Natural Human-Machine Communication. *In the Proceedings of IEEE*. 2000. 88(8):1142-1165.

Juang, B. H. Maximum Likelihood Estimation for Mixture Multi-Variate Stochastic Observations of Markov Chains. *AT&T Technical Journal*. 1985. 3(3): 239-251.

Kirschning, I., Tomabechi, H. and Aoe, J-I. A Parallel Recurrent Cascade-Correlation Neural Network with Natural Connectionist Glue. *Proceedings of* *the 1995 IEEE International Conference on Neural Networks, ICNN*. 1995. 2: 953-956.

Kirschning, I., Tomabechi, H. and Aoe, J-I. Recent Advances in Continuous Speech Recognition Using the Time-Sliced Paradigm. *International Workshop on Soft Computing in Industry 1996, IWSCI'96.* April 1996. Muroran, Japan. 1996.

Kohavi, R. Study of Cross-Validation and Bootstrap Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*. 1995. 2: 1137-1143.

Ladefodged, P. *Phonetic data analysis: An Introduction to Fieldwork and Instrumental Techniques*. 2nd Edn. United Kingdom: Blackwell Publishing. 2003.

Lee, K. F. Context-Dependent Phonetic Hidden Markov Models for Speaker-Independent Continuous Speech Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1990. 38(4): 599-609.

Lee, T., Ching, P. C. and Chan, L. W. Isolated Word Recognition using Modular Recurrent Neural Networks. *Pattern Recognition*. 1998. 31(6): 751-760.

Lippmann, R. P. An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*. 1987. 4(2): 4-22.

Lippmann, R. P. Pattern Classification using Neural Networks. *IEEE Communications Magazine*. 1989. 27(11): 47-50, 59-64.

Li, F. Contrast and Covert Contrast: The Phonetic Development of Voiceless Sibilant Fricative in English and Japanese Toddlers. *Journal of Phonetics*. 2009. 37: 111-124.

Liu, Y., Lee, Y. C., Chen, H. H. and Sun, G. Z. Speech Recognition using Dynamic Time Warping with Neural Network Trained Templates. *International Joint Conference on Neural Networks*. 1992. 2: 326-331.

Looney, C. G. *Pattern Recognition using Neural Networks – Theory and Algorithms for Engineers and Scientists*. New York: Oxford University Press. 1997.

Maaly, I. A. and El-Obaid, M. Speech Recognition using Artificial Neural Networks. *2nd Information and Communication Technologies, ICTTA '06*. 2006. 1: 1246-1247.

Macias-Guarasa, J. Montero, J. M., Ferreiros, J., Cordoba, R., San-Segundo, R., Gutierrez-Arriola, J., D'Haro, L. F., Fernandez, F., Barra, R. and Pardo, J. M. Novel Applications of Neural Networks in Speech Technology Systems: Search Space Reduction and Prosodic Modeling. *Intelligent Automation and Soft Computing*. 2009. 15(4): 631-646.

Matsuura, Y., Miyazawa, H. and Skinner, T. E. Word Recognition using a Neural Network and a Phonetically based DTW. *Proceedings of the 1994 IEEE Workshop on Neural Networks for Signal Processing [1994] 4*. Sept 6-8, 1994. Ermioni, Greece: IEEE. 1994. 329-334.

McLoughlin, I. *Applied Speech and Audio Processing with Matlab Examples*. New York: Cambridge University Press. 2009.

Meyer, A. Estimation of Auditory Spectro-Temporal Receptive Fields using Statistical Learning Methods. Master Thesis. University Oldenburg; 2009.

Moller, M. F. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks*. 1993. 6: 525-533.

Mitra, S. K. *Digital Signal Processing: A Computer-based Approach*. 3rd Edn. Singapore: Mc Graw Hill. 2006.

Nagata, K., Kato, Y. and Chiba, S. Spoken Digit Recognizer for Japanese Language. *Journal of the Audio Engineering Society*, *JAES*. 12(4): 1964.

Nilsson, J. J. *Learning Machines*. New York: McGraw Hill. 1965.

Olson, H. F. and Belar, H. Phonetic Typewriter. *Journal of the Acoustical Society of America*. 1956. 28(6): 1072-1081.

Pallet, D. S. The Role of the National Institute of Standards and Technology in DARPA's Broadcast News Continuous Speech Recognition Research Program. *Speech Communication*. 2002. 37(1-2): 3-14.

Patterson, R. D., Smith, D. R. R., Dinther, R. V. and Walters, T. C. Size Information in the Production and Perception of Communication Sounds. In: Yost, W. A., Popper, A. N. and Fay, R. R. *Auditory Perception of Sound Sources*. New York: Springer Science + Business Media, LLC. 43-75; 2008.

Pisarn, C. and Theeramunkong, T. Thai Spelling System for Automatic Spelling Speech Recognition. *An International Journal on Information Sciences*. 2007. 178: 122-136.

Priddy, K. L. and Keller, P. E. *Artificial Neural Networks – An Introduction*. India: SPIE. 2007.

Rabiner, L. and Juang, B. H. *Fundamentals of Speech Recognition*. New Jersey: Prentice Hall, Englewood Cliffs. 1993.

Rabiner, L. R., Levinson, S. E., Rosenberg, A. E. and Wilpon, J. G. Speaker Independent Recognition of Isolated Words using Clustering Techniques. *IEEE Transactions Acoustics, Speech, and Signal Processing*. 1979. ASSP-27: 336-349.

Radin, G. The 801 Minicomputer. *IBM Journal of Research and Development*. 1983. 27(3): 237-246.

Refaeilzadeh, P., Tang, L. and Liu, H. *Encyclopedia of Database Systems - Cross Validation*. New York: Springer Science+Business Media. 2009.

Rumelhart, D. E. and McClelland, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: M.I.T. Press. 1986.

Sakai, T. and Doshita, S. The Phonetic Typewriter. *Proceedings of International Federation of Information Processing, IFIP Congress*. Aug 27 – Sept 1, 1962. Munich, Germany: ACM New York. 1962. 445-450.

Sakoe, H and Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1978. ASSP-26(1): 43-49.

Selouani, S-A. and Caelen, J. Recognition of Arabic Phonetic Features Using Neural Networks and Knowledge-Based System: a Comparative Study. *Proceedings of IEEE International Joint Symposia on Intelligence and Systems 1998*. May 1998. Rockville, Maryland: IEEE Xplore Press. 1998. 404-411.

Shadle, C. H. and Mair, S. J. Quantifying Spectral Characteristics of Fricatives. *Proceedings of ICSLP '96*. 1996. 3: 1521-1524.

Shadle, C. H., Mair, S. J. Carter J. N. and Millner, N. The Effect of Vowel Context on Acoustic Characteristics of [ç, x]. *Proceedings of the 13th International Congress of Phonetic Sciences, ICPhS '95*. 1995. 1: 66-69.

Sima, M., Croitoru, V. and Burileanu, D. Performance Analysis on Speech Recognition using Neural Networks. *Proceeding of the International Conference on Development & Application Systems*. May 1998. Suceava, Romania: University of Suceava. 1998, 259-266.

Smith, S.W. The Scientist and Engineer's Guide to Digital Signal Processing. California: California Technical Publishing. 1997.

St George, B. A., Wooten, E. C. and Sellami, L. Speech Coding and Phoneme Classification using Matlab and Neural Networks. *Information Sciences Elsevier*, 1997. 90: 109-119.

Steinberg, R. and O'Shaughnessy, D. Segmentation of a Speech Spectrogram Using Mathematical Morphology. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008*. March 31-April 4, 2008. Las Vegas, Nevada, USA: IEEE. 2008. 1637 − 1640.

Sutat, S. T. and Chularat, T. Feature Windowing-Based for Thai Text-Dependent Speaker Identification using MLP with Backpropagation Algorithm. *Proceedings of the 2000 IEEE International Symposium on Circuits and Systems, ISCAS2000*. 2000. 3(3): 579-582.

Suzuki, J. and Nakata, K. Recognition of Japanese Vowels – Preliminary to the Recognition of Speech. *Journal of the Radio Research Laboratories*. 1961. 37(8): 193-212.

Tan, C.L. and Jantan, A. Digit Recognition using Neural Networks. *Malaysian Journal of Computer Science*. 2004. 17(2): 40-54.

Ting, H. N. and Yunus, J. Speaker-Independent Malay Vowel Recognition of Children using Multi-Layer Perceptron. *TENCON 2004*. 2004. A: 68-71 vol 1.

Ting Hua Nong. *Speech Analysis and Classification using Neural Networks for Computer-Based Malay Speech Therapy*. Master Thesis. Universiti Teknologi Malaysia; 2002.

Watrous, R. L. Speaker Normalization and Adaptation using Second-Order Connectionist Networks. *IEEE Transactions on Neural Networks*. 1993. 4(1): 21-30.

Weinberg, B. and Horii, Y. Acoustic Features of Pharyngeal /s/ Fricatives Produced by Speakers with Cleft Palate. *Cleft Palate Journal 12*. 1975. 12(1): 12-16.

Woodland, P. C., Odell, J. J., Valtchev, V., and Young, S. J. Large Vocabulary Continuous Speech Recognition using HTK. *Proceedings of ICASSP '94*. 1994. 1: 251-254.

Wright, J. L. Neural Network Architecture Selection Analysis with Application to Cryptography Location. *WCCI 2010 IEEE World Congress on Computational Intelligence*. July 18-23, 2010. Barcelona, Spain: IEEE. 2010. 2941-2946.

Yeou, M. and Maeda, S. Pharyngeal and Uvular Consonants are Approximants: An Acoustic Modeling Study. *Volume 2 of Proceedings of the XIIIth Intemational Congress of Phonetic Sciences, ICPhS 95*. August 13-19, 1995. Stockholm, Sweden: Kungliga Tekniska Högskolan. 1995. 586-589.

Yusof, Z. M. and Ahmed, M. Malay Phoneme Classification using Perceptual Linear Prediction (PLP) algorithm. *2nd International Conference on Engineering Technology 2009*. 8 – 10 Dec 2009. Kuala Lumpur, Malaysia: UniKL BMI. 2009, 1-5.

Yoshua, B. *Neural Networks for Speech and Sequence Recognition*. London, UK: International Thomson Computer Press. 1996.

Zue, V. W. and Cole, R. A. Experiments on Spectrogram Reading. *IEEE International Conference on Acoustic, Speech, and Signal Processing, ICASSP '79*. 1979. 4: 116-119.

Zue, V. W. *et al*. Jupiter: A Telephone-based Conversational Interface for Weather Information. *IEEE Trans. Speech Audio Processing*. 2000. 8: 85-96.