

PEMBANGUNAN TEKNIK PENENTUAN SELA MASA SENYAP
DALAM SISTEM PENGECAMAN SUARA

AHMAD IDIL BIN ABDUL RAHMAN

UNIVERSITI TEKNOLOGI MALAYSIA

PEMBANGUNAN TEKNIK PENENTUAN SELA MASA SENYAP
DALAM SISTEM PENGECAMAN SUARA

AHMAD IDIL BIN ABDUL RAHMAN

Tesis ini dikemukakan
sebagai memenuhi syarat penganugerahan
ijazah Sarjana Kejuruteraan (Elektrik)

Fakulti Kejuruteraan Elektrik
Universiti Teknologi Malaysia

JULAI 2005

Untuk mak dan abah yang tersayang serta isteri yang tercinta atas dorongan dan sokongan selama ini.

Untuk anakanda yang bakal lahir agar menjadi inspirasi di masa depan.

PENGHARGAAN

Penulis ingin merakamkan penghargaan ikhlas serta jutaan terima kasih kepada penyelia tesis, Prof. Ir. Dr. Sheikh Hussain b. Shaikh Salleh atas bimbingan dan dorongan yang diberi sepanjang tempoh penyelidikan dan penulisan tesis ini. Ucapan terima kasih juga ditujukan kepada Prof. Madya Dr. Ahmad Zuri b. Sha'ameri.

Kerjasama dan bantuan biasiswa daripada Sekolah Pengajian Siswazah Universiti Teknologi Malaysia amatlah dihargai.

Akhir sekali, penghargaan juga ditujukan kepada semua yang terlibat sama ada secara langsung atau tidak langsung dalam membantu menjayakan projek penyelidikan ini.

ABSTRAK

Mengklasifikasikan isyarat suara kepada bahagian 'voiced', 'unvoiced' dan senyap (V/UV/S) merupakan proses yang penting dalam kebanyakan aplikasi pemprosesan isyarat suara seperti sintesis suara, segmentasi dan pengecaman isyarat suara. Dua kaedah pengukuran yang boleh mencerap isyarat 'voiced'/'unvoiced' atau senyap akan cuba dilihat di dalam penyelidikan ini. Kaedah itu ialah '*Instantaneous Energy*' (IE) dan '*Local Time Correlation*' (LTC). Kaedah IE dan LTC adalah antara kaedah terkini untuk analisis isyarat yang tidak tetap dan telah berjaya diaplikasikan dalam pemprosesan isyarat suara. Satu kajian perbandingan akan dibuat menggunakan dua algoritma ini bagi mengklasifikasikan segmen isyarat suara kepada dua kelas: 'voiced'/'unvoiced' dan senyap. Dalam kajian awal, kaedah IE dan LTC akan digunakan untuk mengenalpasti dan membuang sela masa senyap dalam sampel isyarat suara. Eksperimen dilakukan dengan menggunakan '*Linear Predictive Coding*' (LPC) dan '*Dynamic Time Warping*' (DTW) untuk pengecaman digit berasingan dalam Bahasa Malaysia. Teknik tanpa membuang sela senyap LPC-DTW memberikan ketepatan pengecaman 98.28%. Dengan mengenalpasti dan membuang sela senyap, kedua-dua teknik IE-LPC-DTW dan LTC-LPC-DTW memberikan ketepatan pengecaman 98%. Sistem ini kemudiannya diaplikasikan kepada melatih dan menguji pengecaman digit bersambung. Segmentasi masukan digit bersambung dilakukan dengan menggunakan teknik IE dan LTC. Pengecaman digit bersambung menggunakan teknik IE-LPC-DTW memberikan ketepatan pengecaman digit sebanyak 93.3% dan ketepatan pengecaman rentetan digit sebanyak 78%. Tetapi menggunakan teknik LTC-LPC-DTW ketepatan pengecaman digit adalah sebanyak 93.2% dan ketepatan pengecaman rentetan digit sebanyak 77.7%.

ABSTRACT

Classification of speech into voiced, unvoiced and silence (V/UV/S) regions is an important process in many speech processing applications such as speech synthesis, segmentation and speech recognition system. Two such measures are investigated with respect to their ability to discern voiced/unvoiced and silence segments of speech. They are the Instantaneous Energy (IE) and Local Time Correlation (LTC) method. Both IE and LTC methods are recently proposed technique for nonstationary signal analysis and have been successfully applied to speech processing. A comparative study was made using these two algorithms for classifying a given speech segment into two classes: voiced/unvoiced speech and silence. IE and LTC methods were proposed to remove all the silent intervals in speech sample. Experiment are carried out using Linear Predictive Coding (LPC) and Dynamic Time Warping (DTW) for isolated digit recognition in Bahasa Malaysia. The technique without silent removal LPC-DTW gives a recognition accuracy of 98.28%. With detection and removing of silent interval, both technique IE-LPC-DTW and LTC-LPC-DTW gives a recognition accuracy of 98%. The system then are applied for training and testing for connected digit recognition. The segmentation of input string of the digits are carried out using IE and LTC techniques. Connected digit recognition using IE-LPC-DTW had 93.3% digit accuracy and 78% digit string. However using LTC-LPC-DTW the performance decreased to 93.2% and 77.7% respectively.

KANDUNGAN

BAB	TAJUK	MUKA SURAT
	JUDUL	i
	PENGAKUAN	ii
	DEDIKASI	iii
	PENGHARGAAN	iv
	ABSTRAK	v
	ABSTRACT	vi
	KANDUNGAN	vii
	SENARAI JADUAL	xi
	SENARAI RAJAH	xiii
	SENARAI SIMBOL / SINGKATAN	xvi
1	Pengenalan	1
	1.1 Latar Belakang	1
	1.2 Definasi Masalah	3
	1.3 Susunan Tesis	4
2	PEMROSESAN ISYARAT SUARA	6
	2.1 Pengenalan	6
	2.2 Latar Belakang Sistem Pengecaman Pertuturan	7
	2.3 Unit Perwakilan	7
	2.4 Sukukata Dalam Bahasa Malaysia	9

2.5	Rekabentuk Sistem Pengecaman Pertuturan.	12
2.6	Model Pengecaman Pertuturan	14
2.6.1	Penyarian Sifat	15
2.6.2	Pengekodan Ramalan Linar (LPC)	18
2.7	Rumusan	23
3	SEGMENTASI ISYARAT SUARA	24
3.1	Pengenalan	24
3.2	Algoritma Pengesanan Sela Senyap	25
3.3	Kaedah ' <i>Instantaneous Energy</i> '	27
3.4	Kaedah ' <i>Local Time Correlation</i> '	29
4	TEKNIK PENGECEMAN PERTUTURAN	33
4.1	Pengenalan	33
4.2	Pengecaman Ucapan Digit	34
4.3	Teknologi Pengecaman Pertuturan	36
4.3.1	Rangkaian Neural Untuk Pengecaman Suara	37
4.3.2	' <i>Hidden Markov Model</i> ' Untuk Pengecaman Suara	40
4.3.3	' <i>Dynamic Time Warping</i> ' Untuk Pengecaman Suara	42
4.4	Teknik Pengecaman Suara Berasaskan DTW	46
4.5	Rekabentuk DTW	51

5	EKSPERIMEN DAN KEPUTUSAN	53
5.1	Pengenalan	53
5.2	Pengecaman Digit Berasingan Menggunakan Model LPC-DTW	54
5.2.1	Pencapaian dan Penilaian Model LPC-DTW	58
5.3	Pengecaman Digit Berasingan Menggunakan Model IE-LPC-DTW dan LTC-LPC-DTW	62
5.3.1	Pencapaian dan Penilaian Model IE-LPC-DTW	64
5.3.2	Pencapaian dan Penilaian Model LTC-LPC-DTW	67
5.3.3	Rumusan perbandingan antara Model LPC-DTW, IE- LPC-DTW dan LTC-LPC- DTW untuk pengecaman digit terasing	70
5.4	Pengecaman Digit Bersambung Menggunakan Model IE-LPC-DTW dan LTC-LPC-DTW	72
5.4.1	Pencapaian dan Penilaian Model IE-LPC-DTW	78
5.4.2	Pencapaian dan Penilaian Model LTC-LPC-DTW	83

6	KESIMPULAN DAN CADANGAN	86
6.1	Ringkasan Kesimpulan	86
6.2	Cadangan	88
	RUJUKAN	89

SENARAI JADUAL

NO. JADUAL	TAJUK	MUKA SURAT
2.1	Huruf vokal dan penggunaannya	9
2.2	Huruf diftong dan penggunaannya	9
2.3	Huruf fonem dan penggunaannya	10
2.4	Huruf gabungan dan penggunaannya	11
5.1	Penilaian pencapaian untuk set ujian 750 digit bagi model LPC-DTW	60
5.2	Masa yang diambil model LPC-DTW untuk latihan dan ujian	61
5.3	Penilaian pencapaian untuk set ujian 750 digit bagi model IE-LPC-DTW	65
5.4	Masa yang diambil model IE-LPC-DTW untuk latihan dan pengujian	66
5.5	Penilaian pencapaian untuk set ujian 750 digit bagi model LTC-LPC-DTW	68

5.6	Perbandingan ketepatan pengecaman ketiga-tiga model	69
5.7	Masa yang diambil model LTC-LPC-DTW untuk latihan dan ujian	70
5.8	Perbandingan Pencapaian Ketiga-tiga model	71
5.9	Purata Saiz 'Frame' Keseluruhan Sampel	71
5.10	Senarai 30 sebutan digit bersambung untuk proses pengujian pengecaman digit bersambung	73
5.11	Senarai 12 sebutan digit bersambung untuk proses latihan pengecaman digit bersambung	77
5.12	Penilaian pencapaian untuk set ujian 300 rentetan digit bagi model IE-LPC-DTW	80
5.13	Prestasi pengecaman digit bersambung model IE-LPC-DTW	82
5.14	Penilaian pencapaian untuk set ujian 300 rentetan digit bagi model LTC-LPC-DTW	84
5.15	Prestasi pengecaman digit bersambung model LTC-LPC-DTW	85
6.1	Ringkasan prestasi pengecaman ucapan digit terasing	87
6.2	Ringkasan prestasi pengecaman ucapan digit bersambung	87

SENARAI RAJAH

NO. RAJAH	TAJUK	MUKA SURAT
2.1	Proses pengecaman pertuturan menggunakan pencontoh rujukan	12
2.2	Proses pengecaman suara dengan pilihan klasifikasi jujukan. V untuk 'voiced', UV untuk 'unvoiced', P untuk 'plosive' dan F untuk 'fricative'.	13
2.3	Prinsip pengecam suara oleh Wiren dan Stubbs	13
2.4	Gambarajah Blok Asas Pengecam Pertuturan	14
2.5	Perwakilan digital isyarat suara	19
4.1	Neuron Biologi	37
4.2	Neuron Buatan	38
4.3	'Multi-Layer Perceptron'	38
4.4	Ilustrasi HMM dari kiri ke kanan	40
4.5	Laluan 'Dynamic Time Warping'	43

4.6	Ilustrasi laluan penjajaran masa antara dua perkataan yang berbeza dari segi skala masa	46
4.7	Padanan tidak linear dalam algoritma DTW	47
4.8	Had laluan tempatan	49
4.9	Penjajaran laluan ' <i>dynamic programming</i> ' antara dua digit	50
4.10	Cartalir umum DTW	52
5.1	Rajah blok proses latihan sistem pengecaman digit berasingan	54
5.2	Cartalir pengesanan titik mula	56
5.3	Rajah blok proses pengujian sistem pengecaman digit berasingan	57
5.4	Pengecaman digit berasingan menggunakan model IE-LPC-DTW dan LTC-LPC-DTW	62
5.5	Fungsi LTC dan Fungsi IE untuk digit satu	63
5.6	Isyarat suara digit satu	63
5.7	Pengecaman digit bersambung menggunakan model IE-LPC-DTW dan LTC-LPC-DTW	72

5.8	Sampel isyarat suara digit bersambung (410)	74
5.9	Cartalir segmentasi sebutan digit bersambung	75
5.10	Rajah blok proses latihan pengecaman digit bersambung	76
5.11	Rentetan digit '084' yang memberikan ralat penyisipan	81
5.12	Rentetan digit '786' yang memberikan ralat pemotongan	81

SENARAI SIMBOL / SINGKATAN

ASR	-	Automatic Speech Recognition
ANN	-	Artificial Neural Network
DTW	-	Dynamic Time Warping
FFT	-	Fast Fourier Transform
HMM	-	Hidden Markov Model
IE	-	Instantaneous Energy
LAR	-	Log Area Ratio
LPC	-	Linear Predictive Coding
LTC	-	Local Time Correlation
VQ	-	Vector Quantization
ZCR	-	Zero Crossing Rate

BAB 1

PENGENALAN

1.1 Latarbelakang

Bertutur adalah satu cara termudah bagi manusia untuk berhubung antara satu sama lain. Dalam menuju ke era yang lebih efisien, manusia sebenarnya lebih lazim dan selesa dengan pertuturan. Kaedah-kaedah perhubungan yang lain misalnya bahasa isyarat memerlukan lebih tumpuan, pergerakan yang terbatas dan kadang-kala boleh menyebabkan ketegangan disebabkan oleh keadaan yang tidak tabii.

Dalam tahun 1950 an, kebanyakan sistem komputer menggunakan suis sebagai masukan dan membaca keluaran dengan menggunakan LED (*light emitting diode*). Tidak lama selepas itu kad tebuk pula digunakan. Lewat tahun 1970 an terminal CRT (*cathode-ray tube*) pula mengambil tempat dengan menjadikan proses masukan dan keluaran menjadi lebih berkesan. Masukan melalui papan kekunci adalah jauh lebih mudah berbanding suis atau kad tebuk dan membaca huruf lebih cepat daripada menterjemah kod perduaan pada LED atau kad tebuk. Walaubagaimanapun menaip pada papan kekunci adalah lebih lambat berbanding bertutur secara berterusan. Lebih-lebih lagi jika sedang menaip atau membaca, pengguna haruslah memfokuskan diri mereka kepada tugas untuk masukan dan

keluaran itu. Berbeza dengan kaedah pertuturan, pengguna akan lebih mudah untuk melaksanakan tugas tersebut secara bebas.

Melalui pertuturan spontan, manusia dianggarkan dapat berkomunikasi antara 2.0 hingga 3.6 perkataan sesaat (Turn, 1974). Jurutaip yang mahir pula dapat menaip lebih kurang 1.6 hingga 2.5 perkataan sesaat bagi teks yang telah tersedia. Untuk menaip secara spontan atau menyelesaikan masalah, seorang jurutaip yang mahir dapat menaip kira-kira 0.3 perkataan sesaat. Bagi seorang jurutaip yang tidak mahir hanya dapat menaip 0.2 hingga 0.4 perkataan sesaat di bawah keadaan optimum (Newell, 1973). Anggaran kepantasan menulis teks dengan tangan adalah 0.4 perkataan sesaat (Lea, 1980). Membaca secara senyap dapat mencapai 2.5 hingga 9.8 perkataan sesaat, tetapi pembaca mestilah menumpukan perhatian terhadap bacaannya sahaja (Newell, 1973). Maklumat di atas menunjukkan bahawa perantaramuka antara mesin dan manusia adalah lebih optimum melalui kaedah pertuturan.

Dengan keadaan semasa di mana komputer lebih diperlukan di dalam perniagaan, pentadbiran dan pendidikan, adalah perlu kepada perantara muka antara manusia dan mesin yang lebih berkesan dan pantas. Kebanyakan masa pemprosesan komputer digunakan untuk memproses perkataan, kemasukan data dan sebagainya. Dengan membenarkan manusia berkomunikasi dalam suasana pertuturan yang semulajadi, keberkesanan serta kualiti dan hasilnya akan meningkat. Pertuturan juga menawarkan komunikasi yang mudah dan menjimatkan kos untuk jarak yang jauh dan lebih berkesan untuk mereka yang kurang upaya.

Faktor-faktor inilah yang mendorong ke arah penyelidikan dan pembangunan bidang pengecaman pertuturan. Bagaimanapun pencapaian pengecaman merupakan sesuatu yang kompleks dan menjadi tugas yang bukan mudah. Manusia mungkin dapat mempelajari bahasa dengan mudah seperti kanak-kanak yang belajar melalui kaedah pendedahan tetapi mesin memerlukan sistem yang kompleks walaupun hanya untuk melaksanakan tugas pengecaman yang paling asas.

1.2 Definasi Masalah

Sistem pengecaman suara banyak dibangunkan dengan menggunakan digit sebagai unit pengecaman. Ini adalah disebabkan aplikasinya yang meluas dan mudah dikomersilkan seperti kemasukan data inventori yang besar, pengesahan kad atm, panggilan telefon menggunakan suara dan sebagainya. Pembangunan pengecaman digit dimulakan dengan pengecaman ucapan digit secara terasing (*isolated*). Kemudian penyelidikan dikembangkan lagi kepada pengecaman ucapan digit secara bersambung (*connected*). Kaedah pengecaman yang kedua lebih efisien dan mudah diaplikasikan kerana pengecaman secara ucapan terasing bukan sahaja membuatkan pengguna kekok bahkan melambatkan proses kemasukan data atau isyarat suara. Walaubagaimanapun teknik pengecaman ucapan digit secara bersambung mempunyai cabarannya yang tersendiri. Salah satu daripada cabarannya ialah bagaimana untuk mensegmentasikan masukan ucapan digit secara bersambung itu agar proses pengecaman boleh dibuat dengan mudah.

Dalam analisis isyarat suara, teknik untuk mensegmenkan isyarat suara kepada ‘voiced’, ‘unvoiced’ dan senyap telah banyak dibangunkan. Teknik-teknik yang dibangunkan itu bukan sahaja digunakan di dalam pengecaman suara malahan telah diimplementasikan dalam bidang pemprosesan suara yang lain seperti pengkodan suara dan sintesis suara. Dalam proses segmentasi isyarat suara, maklumat seperti sela masa senyap boleh dijadikan sebagai rujukan bagi menentukan sempadan antara digit-digit yang terdapat dalam pengecaman digit bersambung.

Dalam tesis ini kajian yang menggunakan dua teknik untuk penentuan sela masa senyap telah dilakukan. Dua teknik tersebut adalah ‘*Instantaneous Energy*’ (IE) dan ‘*Local Time Correlation*’ (LTC). Untuk menguji keberkesanan dua teknik ini, dua eksperimen iaitu sistem pengecaman ucapan digit terasing dan sistem pengecaman ucapan digit bersambung telah dilakukan. Sistem pengecaman digit ini menggunakan kaedah ‘*Linear Predictive Coding*’ (LPC) untuk proses penyarian

sifat dan kaedah '*Dynamic Time Warping*' (DTW) untuk proses pengecaman isyarat digit.

Untuk pengecaman ucapan digit terasing, teknik IE dan LTC akan diguna bagi mengesan sela masa senyap dalam sebutan digit oleh penutur. Sela masa senyap yang dikesan itu akan dibuang semasa fasa latihan dan fasa pengujian. Perbandingan akan dibuat di antara tiga model pengecaman iaitu LPC-DTW, IE-LPC-DTW dan LTC-LPC-DTW.

Untuk pengecaman ucapan digit bersambung pula, teknik IE dan LTC akan digunapakai bagi tujuan segmentasi masukan rentetan digit. Segmentasi akan dilaksanakan kepada kedua-dua fasa iaitu latihan dan pengujian. Semasa fasa latihan digit-digit yang telah disegmenkan itu akan disimpan sebagai pencontoh (*template*) dalam bentuk digit terasing. Sekali lagi perbandingan dibuat di antara dua model pengecaman iaitu IE-LPC-DTW dan LTC-LPC-DTW.

1.3 Susunan Tesis

Dalam bab 2 dimuatkan tentang kajian pemprosesan isyarat suara (*front end*) yang meliputi latar belakang sistem pengecaman pertuturan, unit perwakilan, rekabentuk serta model pengecaman pertuturan. Turut dimuatkan ialah mengenai teknik-teknik untuk proses penyarian sifat (*feature extraction*) dan huraian khusus dibuat untuk proses penyarian sifat menggunakan teknik '*Linear Predictive Coding*' (LPC).

Dalam bab 3 pula penerangan dimulakan dengan pengenalan klasifikasi '*voiced speech*', '*unvoiced speech*' atau sela senyap (*silence*). Seterusnya

perbincangan dibuat mengenai algoritma yang biasa digunakan untuk penentuan sela masa senyap dalam isyarat suara. Kaedah yang digunakan di dalam tesis ini iaitu '*Instantaneous Energy*' (IE) dan '*Local Time Correlation*' (LTC) juga dibincangkan di akhir bab ini.

Manakala dalam bab 4 pula tumpuan akan diberikan kepada perbincangan mengenai teknik-teknik pengecaman suara. Ulasan dibuat mengenai teknik-teknik pengecaman suara menggunakan '*Neural Network*' (NN), '*Hidden Markov Model*' (HMM) dan '*Dynamic Time Warping*' (DTW).

Bab 5 pula akan menghuraikan tentang metodologi pelaksanaan dan prestasi pencapaian untuk sistem pengecaman ucapan digit terasing dan sistem pengecaman ucapan digit bersambung.

Bab 6 mengandungi kesimpulan dan cadangan untuk kajian ini.

RUJUKAN

- Abdul Alim, O.A., Elboghdadly, N., El Shaar, N.M. (2001). "HMM/NN Hybrids for Continuous Speech Recognition." *Proceeding of National Radio Science Conference NRSC. 2*, 509-516.
- Abdulla, W.H., Chow, D., Sin, G. (2003). "Cross-words Reference Template for DTW-based Speech Recognition Systems." *Proc. of TENCON*, 1576-1579.
- Ainsworth W.A. (1988). "Speech Recognition by Machine." Peter Peregrinus Ltd.
- Alotaibi, Y.A. (2003). "High Performance Arabic Digits Recognizer Using Neural Networks." 670-674.
- Atal, B. S. and Rabiner, L. R. (1976). "A Pattern Recognition Approach to Voice-Unvoiced-Silence Classification with Applications to Speech Recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing. 24*, 201-212.
- Atal, B.S. (1976), "Automatic Recognition of Speakers From Their Voices." *Proceedings of The IEEE. 64*, 460-474.
- Ben-Yishai, A., Burshtein, D. (2004). "A Discriminative Training Algorithm for Hidden Markov Models." *IEEE Trans. On Speech and Audio Processing. 12*, 204-217.
- Beritelli, F., Casale, S., Serrano, S. (2002). "A Robust Speaker Dependent Algorithm for Isolated Word Recognition." *14th International Conference. 2*, 993-996.

- Bridle J.S. and Brown M.D. (1979). "Connected Word Recognition Using Whole Word Templates." *Proc. Autumn Conf. Institute of Acoustics*, 263-265.
- Brigham, E. O. (1974). "The Fast Fourier Transform." Englewood Cliffs: Prentice Hall.
- Casarotto, S., Cerutti, S., Bianchi, A.M. (2003). "Dynamic Time Warping in The Study of ERPs in Dyslexic Children." *Proc. International Conference of The IEEE EMBS*, 2311-2314.
- Cernys, P., Kubilius, V., Macerauskas, V. (2003). " Intelligent Control of The Lift Model." *IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application*: 428-431.
- Cheng, X., Wang, H., Li, Z. (2002). "Speech Adaptation Using Neural Networks for Connected Digit Recognition." *Proc. Of International Conference on Neural Information Processing*. 5, 2401-2404.
- Deller, J.R, Hansen, J.H.L. (2000). "Discrete-Time Processing of Speech Signals." IEEE Press Editorial Board.
- Denes, P. and Mathew, M. V. (1960). "Spoken Digit Recognition Using Time-Frequency Pattern Matching." *J. Acoust. Soc. Am.*, 32, 1450-1455.
- Dumitru, C. O., Gavat, I. (2003). " Voice-Dial by Statistical Recognition of Continuous Speech.": 157-160.
- Flaherty, M.J., Sidney, T. (1994). " Real Time Implementation of HMM Speech Recognition for Telecommunications Applications." *IEEE International Conference 6*, VI/145 - VI/148.

- Gadallah, M., Soleit, E., Mahran, A. (1999). "Noise Immune Speech Recognition System." *National Radio Science Conference NSRC*, C21/1 – C21/8.
- Gan, C. K. and Donaldson, R. W. (1998). "Adaptive Silence Deletion for Speech Storage and Voice Mail Applications." *IEEE Transactions On Acoustics, Speech, And Signal Processing*. 36, 924-927
- Hai, J., Joo, E. M. (2003). "Improved Linear Predictive Coding Method for Speech Recognition" , *ICICS-PCM*. 1614-1618.
- Hermansky, H., & Morgan, N. (1994). "RASTA processing of speech." *Trans. Speech & Audio Proc.* 2, 578-589.
- Huang, B., Kinsner, W. (2002). "ECG Frame Classification Using Dynamic Time Warping." *Proc. Of IEEE Canadian Conference on Electrical & Computer Engineering*, 1105-1110.
- Huang, X., Jack, M. (1988). " On Several Problems of Hidden Markov Models." *Proc. Speech 7th FASE Symposium*, 17-22.
- Ismail Dahaman (1996). "Pedoman Ejaan dan Sebutan Bahasa Melayu." Dewan Bahasa dan Pustaka.
- Junqua, J.C., Wakita, H. (1993). "Evaluation and Optimization of Perceptually-Based ASR Front End." *IEEE Trans. Speech & Audio Proc.* 1, 39-48.
- Lang, K., Waibel, A., and Hinton, G. (1990). "A Time-Delay Neural Network Architecture for Isolated Word Recognition". *Neural Network* 3(1): 23-43.
- Lea, Wayne A. (1980). "Trends in Speech Recognition." Prentice-Hall.

- Li, T. H. and Gibson, J. D. (1996). "Time-correlation Analysis of Nonstationary Signals With Application to Speech Processing." *Proceedings of the IEEE-SP International Symposium*, 449-452.
- Li, T. H. and Gibson, J. D. (1997). "Time-correlation Analysis of A Class of Nonstationary Signals With an Application To Radar Imaging." *IEEE International Conference*, 3765-3768.
- Liao, L. and Gregory, M. A. (1999). "Algorithm for Speech Classification." *Fifth International Symposium on Signal Processing and its Applications*, 623-627.
- Lippmann, R.P (1989). "Review of Neural Networks for Speech Recognition." *Neural Computation*. 1. 1-38.
- Liu, C., Lin, M., Wang, W. (1990). "Study of Line Spectrum Pair Frequencies for Speaker Recognition." *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*. 1, 277-280.
- Loo, C. and Donaldson, R. W. (1997). "An Adaptive Silence Deletion Algorithm for Compression of Telephone Speech." *IEEE Pacific Rim Conference*. 2, 701-705.
- Makhoul, John (1975). "Linear Prediction in Automatic Speech Recognition." *IEEE Symposium on Speech Recognition, Academic Press, London*. 100-103.
- Maragos, P., Loupas, T., Pitsikalis, V. (2002). "On Improving Doppler Ultrasound Spectroscopy With Multiband Instantaneous Energy Separation." *IEEE 14th International Conference*, 611-614.
- Nakatsu, R. and Kohda, M. (1974). "Computer Recognition Of Spoken Connected Words Based On VCV Syllable." *Autumn Meet. Acoust. Soc. Japan*. 255-256.

- Neelakantan, V. and Gowdy, J.N. (1992). "A Comparative Study of Using Different Speech Parameters in The Design of a Discrete Hidden Markov Model."
- Newell, Allen, et al. (1973). "Speech Understandings System: Final Report of a Study Group." North-Holland Publishing.
- Novak, D., Cuesta-Frau, D. (2004). " Speech Recognition Methods Applied to Biomedical Signals Processing." *Proc. Of International Conference of IEEE EMBS*, 118-121.
- Openshaw, J., & Mason, J. (1994). " Optimal Noise-masking of Cepstral Features for Robust Speaker Identification." ESCA Workshop on Automatic Speaker Recognition Identification and Verification. 231-234.
- Parsons, T.W. (1986). "Voice and Speech Processing." McGraw-Hill.
- Peeling, S., & Moore, R. (1988). "Isolated Digit Recognition Using Multilayer Perceptron." *EURASIP Journal Speech Communication*. **7**.
- Picone, J. (1990). "Continuous Speech Recognition Using Hidden Markov Model." In: IEEE ASSP Magazine.
- Rabiner, L. R. and Sambur, M. R. (1977), "Application of an LPC Distance Measure to the Voice-Unvoiced-Silence Detection Problem." *IEEE Transactions on Acoustics, Speech, and Signal Processing*. **25**, 338-343.
- Rabiner, L. R., Juang, B. (1993). "Fundamentals of Speech Recognition." Englewood Cliffs: Prentice Hall.
- Rabiner, L. R., Sambur, M. R. (1976). "Some Preliminary Experiments in the Recognition of Connected Digits" *IEEE Trans. Acoust., Speech and Signal Processing*. ASSP-24, 170-182.

- Rabiner, L. R., Schmidt, C. E. (1980). "Application of Dynamic Time Warping to Connected Digit Recognition." *IEEE Trans. Acoust., Speech and Signal Processing. ASSP-28*, No. 4.
- Rabiner, L., Levinson, S., & Sondhi, M. (1983). "On The Application of Vector Quantization and Hidden Markov Model to Speaker Independent Isolated Word Recognition." In: *The Bell System Technical Journal*. **62**.
- Rath, T.M., Manmatha, R. (2003). "Word Image Matching Using Dynamic Time Warping." *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, 1-7.
- Rose, C. and Donaldson, R. W. (1991). "Real-Time Implementation And Evaluation Of An Adaptive Silence Deletion Algorithm For Speech Compression." *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 461-468.
- Sakoe H. (1979). "Two Level DP Matching – A Dynamic Programming Based Pattern Matching Algorithm For Connected Word Recognition." *IEEE Trans. Acoust., Speech and Signal Processing vol. ASSP-27*, 588-595.
- Sakoe H. and Chiba S. (1971). "A Dynamic Programming Approach To Continuous Speech Recognition." *Proc. 7th ICA*, paper 20C13.
- Sambur, M. (1976). "Speaker Recognition Using Orthogonal Linear Prediction." *IEEE Trans. On Acoustic, Speech, and Signal Processing. ASSP-24 no 4*, 283-289.

- Sang-Hwa, Min-Uk (1999). "A Parallel Phoneme Recognition Algorithm Based on Continuous Hidden Markov Model." *13th International and 10th Symposium on Parallel and Distributed Processing*, 453-457.
- Sheikh Hussain (1993). "A Comparative Study of The Traditional Classifier and The Connectionist Model for Speaker Dependent Speech Recognition System." Universiti Teknologi Malaysia: Tesis Master.
- Sheikh Hussain (1997). "An Evaluation of Preprocessors for Neural Network Speaker Verification." University of Edinburgh: Tesis Ph.D.
- Sheikh Hussain, McJunes F.R and Jack M.A. (1995). "Enhanced Automatic Speaker Verification Based on a Combination of Hidden Markov Model and Multilayer Perceptron." MICC.
- Shuzo Saito, Kazuo Nakata (1985). "Fundamentals of Speech Signal Processing." Academic Press, Inc.
- Sivakumar, S.C., Phillips, W.J., Robertson, W. (2000). "Isolated Digit Recognition Using A Block Diagonal Recurrent Neural Network." *IEEE International Conference*, 726-729.
- Trent, L., Rader, C., & Reynolds, D. (1994). "Using Higher Order Statistic to Increase the Noise Robustness of A Speaker Identification System." ESCA Workshop on Automatic Speaker Recognition Identification and Verification. 221-224.
- Tsuruta S. (1978). "DP-100 Voice Recognition System Achieves High Efficiency." *J. Eng. Educ.*, 50-54.

Waibel, Alex and Kai-Fu Lee (1990) "Reading in Speech Recognition." Morgan Kaufmann.

Wiren, J., and Stubbs H. L. (1956). "Electronic Binary Selection System for Phoneme Classification." *J. Acoust. Soc. Am.* 28. 1082.

Zbancioc, M., Costin, M. (2003). Using Neural Networks and LPCC to Improve Speech Recognition. *IEEE International Conference*, 445-448.

Zelinski, R., Class, F. (1983). "A Segmentation Algorithm for Connected Word Recognition Based on Estimation Principles." *IEEE Trans. Acoust., Speech and Signal Processing. ASSP-31*, 818-827.