AUTOMATIC TRANSCRIPTION AND PHONETIC LABELLING OF DYSLEXIC CHILDREN'S READING IN BAHASA MELAYU

NIK NURHIDAYAT BINTI NIK HIM

SCHOOL OF COMPUTING
UUM COLLEGE OF ARTS AND SCIENCES
UNIVERSITI UTARA MALAYSIA
2015

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to :

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Pengecaman suara automatik *(ASR)* berpotensi untuk membantu kanak-kanak disleksia yang mengalami masalah pembelajaran. Kesalahan dalam penyebutan fonetik yang hampir sama oleh kanak-kanak disleksia amat tinggi sehingga memberi kesan kepada ketepatan pengecaman ASR. Oleh itu, objektif utama kajian ini adalah untuk menilai penerimaan ketepatan ASR dengan menggunakan transkripsi dan pelabelan fonetik automatik untuk kanak-kanak disleksia. Bagi mencapai matlamat utama tersebut, terdapat tiga objektif yang telah ditetapkan: pertama untuk menghasilkan transkripsi dan pelabelan fonetik manual; kedua untuk membina transkripsi dan pelabelan fonetik automatik menggunakan kaedah penjajaran paksa; dan ketiga untuk membandingkan ketepatan di antara transkripsi dan pelabelan fonetik automatik dengan transkripsi dan pelabelan fonetik manual. Lantaran itu, untuk mencapai matlamat kajian ini beberapa kaedah telah digunakan, termasuk pelabelan ucapan dan segmentasi manual, penjajaran paksa, *Hidden Markov Model (HMM)* dan Rangkaian Neural Buatan *(ANN)* untuk proses latihan, dan bagi mengukur ketepatan daripada ASR, Kadar Kesalahan Perkataan (*WER)* dan *False Alarm Rate* (FAR) digunakan. Sebanyak 585 fail ucapan telah digunakan untuk transkripsi manual, penjajaran paksa dan juga proses latihan. Pengecaman yang dijana oleh ASR enjin yang menggunakan transkripsi dan pelabelan fonetik automatik telah mencapai keputusan yang paling optimum iaitu 76.04% dengan kadar WER serendah 23.96% dan FAR 17.9%. Keputusan ini adalah hampir sama dengan ASR enjin yang menggunakan transkripsi dan pelabelan fonetik manual iaitu 76.26%, WER serendah 23.97% dan FAR 17.9%. Kesimpulannya, ketepatan daripada transkripsi dan pelabelan fonetik automatik adalah diterima bagi membantu kanak-kanak disleksia belajar menggunakan ASR dalam Bahasa Melayu (BM).

Kata Kunci: Pembacaan kanak-kanak disleksia, Transkripsi manual, Transkripsi dan pelabelan fonetik automatik, Penjajaran paksa, Pengukuran ketepatan ASR enjin.

Abstract

Automatic speech recognition (ASR) is potentially helpful for children who suffer from dyslexia. Highly phonetically similar errors of dyslexic children's reading affect the accuracy of ASR. Thus, this study aims to evaluate acceptable accuracy of ASR using automatic transcription and phonetic labelling of dyslexic children's reading in BM. For that, three objectives have been set: first to produce manual transcription and phonetic labelling; second to construct automatic transcription and phonetic labelling using forced alignment; and third to compare between accuracy using automatic transcription and phonetic labelling and manual transcription and phonetic labelling. Therefore, to accomplish these goals methods have been used including manual speech labelling and segmentation, forced alignment, Hidden Markov Model (HMM) and Artificial Neural Network (ANN) for training, and for measure accuracy of ASR, Word Error Rate (WER) and False Alarm Rate (FAR) were used. A number of 585 speech files are used for manual transcription, forced alignment and training experiment. The recognition ASR engine using automatic transcription and phonetic labelling obtained optimum results is 76.04% with WER as low as 23.96% and FAR is 17.9%. These results are almost similar with ASR engine using manual transcription namely 76.26%, WER as low as 23.97% and FAR a 17.9%. As conclusion, the accuracy of automatic transcription and phonetic labelling is acceptable to use it for help dyslexic children learning using ASR in Bahasa Melayu (BM).

Keywords: Dyslexic children's reading, Manual transcription, Automatic transcription and phonetic labelling, Forced alignment, Evaluation accuracy of ASR engine.

Acknowledgement

In The Name of ALLAH, Most Gracious, Most Merciful and Big Gratitude to Prophet, Muhammad S.A.W.

First and foremost, I thanked ALLAH the All Mighty for I am blessed to complete this study in time. Special thanks to Dr Husniza Binti Husni, my very helpful, supportive and dedicated supervisor for all her supervision, comments, ideas, suggestion and guideline given to me in order to complete this study.

My special thanks to Dr Mohd Hasbullah Bin Omar and Dr Norliza Katuk for the explanation and guidelines given to me especially during the preparation period and also during the presentation of this study. A special thanks also to all lecturers in Universiti Utara Malaysia for their great help and support during my academic career.

To my beloved family, a million thank you for their moral support and motivation especially my dad, Nik Him Bin Nik Ya and my mom Rohana Binti Kadir. Thanks for the love, encouragement, support and prayers. Last but not least, my fellow friends and others who have contributed directly and indirectly towards the completion of this study.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Network
ART	Automatic Reading Tutor
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition
BM	Bahasa Melayu
C	Consonant
CV	Consonant Vowel
CALL	Computer-assisted language learning
CoLiT	Colorado Literacy Tutor
CSLU	Center for Spoken Language Understanding
FAR	False Alarm Rate
HMM	Hidden Markov Model
HTK	Hidden Markov Modelling Toolkit
IPA	International Phonetic Alphabet
IRT	Interactive Reading Tutor
LD	Learning Disability
MDR	Miscue Detection Rate
MS	Milliseconds
NN	Neural Network
SER	Sentences Error Rate
TTS	Text to Speech
V	Vowel
WER	Word Error Rate

CHAPTER ONE
INTRODUCTION

1.1 Introduction

Automatic speech recognition (ASR) has been an essential technology, and it has come to a stage where it has been actively applied in a lot of industrial and consumer applications. ASR research is still in early stage in Malaysia for Bahasa Melayu (BM). However, ASR can play an important role in the education field like to boost children's is interest in learning. The availability of ASR technology gives opportunity to help children especially dyslexics to enhance their learning ability by using Automatic Reading Tutor (ART) or Interactive Reading Tutor (IRT). In order to develop ART and IRT using ASR technology, speech files of dyslexic children's reading aloud are used to perform transcription and phonetic labelling that serve as important basic elements for the construction of ASR engine (Athanaselis, Bakamidis, Dologlou, Argyriou, & Symvonis, 2014; Taileb, Al-Saggaf, Al-Ghamdi, Al-Zebaidi, & Al-Sahafi, 2013; Pedersen & Larsen, 2010; Husniza & Zulikha, 2009; Li, Deng, Ju, & Acero, 2008; Chuchiarini & Strik, 2003).

Since transcription and phonetic labelling are used for ASR engines, so the training and evaluation accuracy of it must be done by using standard methods and metrics (e.g. hybrids Hidden Markov Model (HMM) and Artificial Neural Network (ANN) for training; Word Error Rate (WER) and False Alarm Rate (FAR) for measuring accuracy). However, in this study the dyslexic children's speech presents a challenge to perform transcription and phonetic labelling due to dealing with highly

The contents of the thesis is for internal user only

REFERENCES

Abushariah, A. A. M., Gunawan, T. S., Khalifa, O. O., Abushariah, M. A. M. (2010). English digits speech recognition system based on Hidden Markov Models. In *International Conference on Computer and Communication Engineering (ICCCE)*, Kuala Lumpur, Malaysia.

Al-Manie, M. A., Alkanhal, M. I., & Al-Ghamdi, M. M. (2009). Automatic speech segmentation using the Arabic phonetic database. In *Proceedings of the World Scientific and Engineering Academy and Society (WSEAS), Automation & Information*, *10*, 6-79.

Athanaselis, T., Bakamidis, S., Dologlou, I., Argyriou, E. N., & Symvonis, A. (2014). Making assistive reading tools user friendly: a new platform for Greek dyslexic students empower by automatic speech recognition. *Multimedia Tools and Application*, *68*(3), 681-699.

Azam, S. M., Mansoor, Z. A., Mughal, M. S., & Mohsin, S. (2007). Urdu spoken digits recognition using classified MFCC and backprogation neural network. In *Computer Graphics, Imaging and Visualisation, IEEE, 7*, 414-418.

Banerjee, S., Beck, J. E., & Mostow, J. (2003). Evaluating the Effect of Predicting Oral Reading Miscues. In *Proceedings of the European Conference on Speech Communication and Technology (Eurospeech), 8*.

Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2000). Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication*, *33*(1), 5-22.

Bauer, T., Hitzenberger, L., & Hennecle, L. (2002). Effects of manual phonetic transcriptions on recognition accuracy of streetnames. In *Proceedings of the International Symposiums for Information Swissenschaft (ISI), 8,* 21-25.

Bhotto, M. Z. A., & Amin, M. R. (2004). Bengali text dependent speaker identification using melfrequency cepstrum coefficient and vector quantization. In *International Conference on Electrical & Computer Engineering (ICECE), 3,* 28-30.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (version 5.4.08) [computer program]. Retrieved April 11, 2015, from http://www.fon.hum.uva.nl/praat/manual/Intro.html.

Bourassa, D., & Treiman, R. (2003). Spelling in children with Dyslexia: Analysis from the Treiman-Bourassa Early spelling test. *Scientific studies of reading, 7*(4), 309-333.

Bourlard, H. A., & Morgan, N. (2012). Connectionist speech recognition: A hybrid approach. *Springer Science & Business Media*, *247.*

Brognaux, S., Roekhaut, S., Drugman, T., & Beaufort, R. (2012). Train & Align: A new online tool for automatic phonetic alignment. In *IEEE Workshop on Spoken Language Technologies,* 416-421.

Cangemi, F., Cutugno, F., Ludusan, B., Seppi, D., & Van C. D. (2011). Automatic Speech Segmentation for Italian (Assi): Tools, Models, Evaluation, and Applications. In *Proceedings of the Associazione Italiana di Scienze della Voce (AISV)*, Lecce, Italy, *7*, 337-344.

Carroll, J. M., & Myers, J. M. (2010). Speech and language difficulties in children with and without a family history of dyslexia. *Scientific Studies of Reading*, *14*(3), 247-265.

Castles, A., Wilson, K., & Coltheart, M. (2011). Early orthographic influences on phonemic awareness tasks: evidence from a preschool training study. *Journal of Experimental Child Psychology*, *108*(1), 203-210.

Chang, S., Shastri, L., & Greenberg, S. (2000). Automatic Phonetic transcription of spontaneous speech (American English). In *Proceedings of the International conferences on Spoken Languages Processing*, Beijing, China, *6*, 330-333.

Chou, F. C., Tseng, C. Y., & Lee, L. S. (2002). A set of corpus-based text-to-speech synthesis technologies for Mandarin Chinese. *Speech and Audio Processing, IEEE Transactions on*, *10*(7), 481-494.

Conn, N., & McTear, M. (2000). Speech Technology: A Solution for People with Disabilities. In *IEEE Seminar on Speech and Language Processing for Disabled and Elderly People*, *7*, 1-6.

Cosi, P., & Hosom, J. P. (1999). Hmm/Neural Network-Based System for Italian Continuous Digit Recognition. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, *14*, 1669-1672.

Choudhary, A., Chauhan, M. R., & Gupta, M. G. (2013). Automatic speech recognition system for isolated & connected words of Hindi language by using Hidden Markov Model Toolkit (HTK). In *Proceedings of the International Conference on Emerging Trends in Engineering and Technology (ACEEE)*, 847-853.

Cucchiarini.C., & Strik, H. (2003). Automatic phonetic transcription: An overview. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, Barcelona, *15*, 347–350.

Das, R., Izak, J., Yuan, J., & Liberman, M. (2010). Forced alignment under adverse conditions. *University of Pennsylvania, CIS Dept*. *Senior Design Project Report.*

DeFries, J. C., Olson, R. K., Pennington, B. F., & Smith, S. D. (1991). Colorado Reading Project: Past, present, and future. *Learning Disabilities: A Multidisciplinary Journal*, *2*, 37-46.

Demuynck, K., & Laureys, T. (2002). A comparison of different approaches to automatic speech segmentation. In *Text, Speech and Dialogue*, *5*, 277-284.

Dinarelli, M., Moschitti, A., & Riccardi, G. (2009). Concept Segmentation and Labeling for Conversational Speech. In *Annual Conference of the International Speech Communication Association*, *10*, 2747-2750.

Douklias, S., Masterson, J., & Hanley, J. R. (2010). Surface and phonological developmental dyslexia in Greek. *Cognitive Neuropsychology*, *26*, 705-723.

Dupuis, A. (2011). Automatic transcription of audio files and why manual transcription may be better. Retrieved March 23, 2015, from: http://www.researchware.com/company/blog/368-automatic transcription.html.

Evermann, G. (1999). Minimum word error rate decoding. *Cambridge University, UK*, 45-67.

Fadhilah, R., & Ainon, R., N. (2008). Isolated Malay speech recognition using Hidden Markov models. *Proceedings of the International Conferences on Computer and Communication Engineering*, 721-725.

Fang, C. (2009). From Dynamic Time Warping (DTW) to Hidden Markov Model (HMM). Final Project report, University of Cincinnati.

Fish, R., Hu, Q., & Boykin, S. (2006). Using audio quality to predict word error rate in an automatic speech recognition system. *Unpublished from MITRE corporation.*

Frikha, M., & Hamida, A. B. (2012). A comparative survey of ANN and hybrid HMM/ANN architectures for robust speech recognition. *American Journal of Intelligent Systems*, *2*(1), 1-8.

Gemello, R., Mana, F., & Albesano, D. (2010). Hybrid HMM/Neural Network based Speech Recognition in Loquendo ASR. Retrieved December, 2, 2014, from http://www. loquendo. com/en/.

Gianna, A., Mclaughlin, T. F., Derby K. M., & Waco, T. (2012). The effects of the Davis symbol mastery system to assist a fourth grader with dyslexia. In *Spelling: A Case Report. I-manager's Journal on Educational Psychology*, *6*(2) 13-18.

Gibbon, D. (1997). Part 1: Spoken language system and corpus design. In *Handbook of standards and resources for spoken language systems*. Berlin: Mouton de Gruyter, 152.

Giurgiu, M., & Kabir, A. (2012). Automatic transcription and speech recognition of Romanian corpus RO-GRID. In *International Conference of the Telecommunications and Signal Processing (TSP), 35*, 465-468.

Goldman, J. P., & Schwab, S. (2014). Easyalign Spanish: An (Semi-) Automatic Segmentation Tool Under Praat. In *Salvador Plans, A. Fonética Experimental, Education Superior Investigation.* Madrid, *1*, 629-640.

Goldman, J. P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. In *Annual Conference of the International Speech Communication Association*, Folorence, *12*, 3233-3236.

Handler, S. M., & Fierson, W. M. (2011). Learning disabilities, dyslexia, and vision. *Paediatrics*, *127*(3), 818-856.

Hagen, A., Pellom, B., & Cole, R. (2003). Children's speech recognition with application to interactive books and tutors. In *Proceedings of the Automatic Speech Recognition and Understanding* (*ASRU), 3*, 186-191.

Hagen, A. (2006). Advances in children's speech recognition with application to interactive literacy tutors. Doctoral dissertation, University of Colorado.

Haykin, S. (1999). Neural networks: a comprehensive foundation. (2nd ed.) Upper Saddle Rever, New Jersey: Prentice Hall.

Hazen, T. J. (2006). Automatic alignment and error correction of human generated transcripts for long speech recordings. *Proceedings of International Conference on Spoken Language Processing*, Pittsburgh, *9*, 1606-1609.

Hieronymus, L. J. (1993). ASCII Phonetic Symbols for the world's Languages: Worldbet, Bell laboratories manuscript.

Hofmann, S., & Pfister, B. (2010). Fully automatic segmentation for prosodic speech corpora. In *Eleventh Annual Conference of the International Speech Communication Association*, Makuhari, Japan, 1389-1392.

Hosom, J. P. (2002). A Comparison of speech recognizers created using manually-aligned and automatically-aligned training data. *Technical Report CSE-00-02,* Oregon Graduate Institute of Science and Technology, Center for spoken Language Understanding, Beaverton.

Hosom, J. P. Shriberg, L., & Green, J. R. (2004). Diagnostic assessment of childhood apraksia of speech using automatic speech recognition (ASR) methods. *Journal of medical speech-language pathology*, *12*(4), 167.

Hosom, O., Villiers, J., Cole, R., Fanty, M., Schalkwyk, J., Yan, Y., & Wei, W. (2006). Training HMM/ANN Hybrids for Automatic Speech Recognition. Retrieved July 3, 2014, from http://www.cslu.ogi.edu/tutordemos/nnet_training/tutorial.html

Hosom, J. P. (2009). Speaker-independent phoneme alignment using transition-dependent states. *Speech Communication, 51*(4), 352-368.

Husniza, H., & Zulikha, J. (2009). Dyslexic children's reading pattern as input for ASR: Data, analysis, and pronunciation model. *Journal of Information and Communication Technology*, *8*, 1-13.

Husniza, H. (2010). Automatic speech recognition model for dyslexic children reading in bahasa Melayu. Doctoral dissertation, Universiti Utara Malaysia.

Husniza, H., & Zulikha, J. (2010). Improving ASR performances using context-dependent phoneme models. *Journal of Systems and Information Technology (JSIT), 12*(1), 56-69.

Husniza, H., Yuhanis, Y., & Siti Sakira, K. (2013a). Speech Malay language influence on automatic transcription and segmentation. *Proceeding of the International Conferences on Computing and Informatics*, ICOCI, Sarawak, Malaysia, *4*, 132-137.

Husniza, H., Yuhanis, Y., & Siti Sakira, K. (2013b). Evaluation of phonetic labeling and segmentation for dyslexic children's speech. *Proceeding of the World Congress one Engineering*, London, U.K, *2*.

Jackson, M. (2005). Automatic Speech Recognition: Human Computer Interface for Kinyarwanda Language. Master dissertation, Computer Science of Makerere University.

Jakovljevic, N., Miskovic, D., Pekar, D., Secujski, M., & Delic, V. (2012). Automatic Phonetic Segmentation for a Speech Corpus of Hebrew, *Infotch-Jahorina, 11*, 742-745.

Jiang, H. (2005). Confidence measures for speech recognition: A survey. *Speech communication*, *45*(4), 455-470.

Jiang, F., Yuan, J., Tsaftaris, S. A., & Katsaggelos, A. K. (2011). Anomalous video event detection using spatiotemporal context. *Computer Vision and Image Understanding, 115*(3), 323-333.

Jurafsky, D., & James, H. (2000). Speech and language processing: *An introduction to natural language processing, computational linguistics, and speech*. Prentice Hall, New Jersey, USA, *2*.

Kabir, A., Barker, J., & Giurgiu, M. (2010). Integrating hidden Markov model and PRAAT: a toolbox for robust automatic speech transcription. In *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, 7745*.

Kaur, E. A., & Singh, E. T. (2010). Segmentation of continuous Punjabi speech signal into syllables. In *Proceedings of the World Congress on Engineering and Computer Science*, *1*, 20-22.

Kawachale, M. S., & Chitode, J. S. (2012). Relative functional comparison of neural and non-neural approaches for syllable segmentation in Devnagari TTS system. *Proceedings of the International Journal of Computer Science Issues (IJCSI)*, *9*(3), 534-543.

Kawai, H., & Toda, T. (2004). An evaluation of automatic phone segmentation for concatenative speech synthesis. In *Proceedings of the International Conference Acoustics, Speech, and Signal Processing (ICASSP'04)*, *1*, 677-680.

Kheir, R., & Way, T. (2006). Improving speech recognition to assist real time classroom note taking. In *Proceedings of Rehabilitation Engineering Society of North America (RESNA) Conference*, *29*, 1-4.

Kim, Y. J., & Gibbon, D. C. (2011). Automatic Learning in Content Indexing Service Using Phonetic Alignment. In *Annual Conference of the International Speech Communication Association*, *12*, 925-928.

Kimball, O., Kao, C. L., Arvizo, T., Makhoul, J., & Iyer, R. (2004). Quick transcription and automatic segmentation of the Fisher conversational telephone speech corpus. In *Proceedings of Rich Transcription Workshop*, Palisades, Newyork.

Kuo, J. W., & Wang, H. M. (2006). A minimum boundary error framework for automatic phonetic segmentation. In *Proceedings of the International Conference on Chinese Spoken Language Processing*. Springer-Verlag, *5*, 399-409.

Kuo, J. W., Lo, H. Y., & Wang, H. M. (2007). Improved HMM/SVM methods for automatic phoneme segmentation. In *Annual Conference of the International Speech Communication Association*, *8*, 2057-2060.

Kvale, K.(1993). Segmentation and Labeling of Speech. (A Dissertation The Doctoral Degree, *The Norwegian Institute of Technology*).

Lakra, S., Prasad, T. V., Sharma, D. K., Atrey, S. H., & Sharma, A. K. (2012). Application of fuzzy mathematics to speech-to-text conversion by elimination of paralinguistic content. In *Proceedings of National Conferences on Soft Computing and Artificial Intelligence, arXiv preprint arXiv:1209.4535*, 294-299.

Lee, C. C., Katsamanis, A., Black, M. P., Baucom, B. R., Georgiou, P. G., & Narayanan, S. S. (2011). Affective state recognition in married couples' interactions using PCA-based vocal entrainment measures with multiple instance learning. In *Proceedings of the International Conferences on Affective Computer Intelligent Interaction (ACII), 2*, 31-41.

Lee, K., Hagen, A., Romanyshyn, N., Martin, S., & Pellom, B. (2004). Analysis and detection of reading miscues for interactive literacy tutors. In *Proceedings of the international conference on Computational Linguistics*. Association for Computational Linguistics. *20*, 1254.

Lee, L. W. (2008). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Annals of Dyslexia*, *58*(1), 37-57.

Leither, C. (2008). Data-Based Automatic Phonetic Transcription. Diploma Thesis, Signal Processing and Speech Communication Lab Graz University of Technology.

Levy, C., Linares, G., Bonastre, J. F., Stepmind, S. A., & Cannet, L. (2005). Mobile phone embedded digit-recognition. In *Workshop on DSP in Mobile and Vehicular Systems,* Sesimbra, Portugal.

Li, X., Ju, Y. C., Deng, L., & Acero, A. (2007). Efficient and robust language modeling in an automatic children's reading tutor system. In *International Conference* on *Acoustics, Speech and Signal Processing (ICASSP), 4*, 193-196.

Li, X., Deng, L., Ju, Y. C., & Acero, A. (2008). Automatic children's reading tutor on hand-held devices. In *Annual Conference of the International Speech Communication Association, 9*, 1733-1736.

Lin, C. Y., Jang, J. S. R., & Chen, K. T. (2005). Automatic segmentation and labeling for Mandarin Chinese speech corpora for concatenation-based TTS. *Computational Linguistics and Chinese Language Processing*, *10*(2), 145-166.

Lu, L., Ghoshal, A., & Renals, S. (2013). Acoustic data-driven pronunciation lexicon for large vocabulary speech recognition. In *IEEE Workshop on Automatic Speech Recognition and Understanding*, 374-379.

Mandal, S., Das, B., Mitra, P., & Basu, A. (2011). Developing Bengali speech corpus for phone recognizer using optimum text selection technique. *International Conference in Asian Language Processing (IALP)*, IEEE Computer Society. 268-271.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. In *Computational linguistics*, *19*(2), 313-330.

Martens, J. P., Binnenpoorte, D., Demuynck, K., Van P. R., Laureys, T., Goedertier, W., et al. (2002). Word Segmentation in the Spoken Dutch Corpus. In *International conference on Language Resources and Evaluation (LREC), 3*, 1432-1437.

McIntyre, C. W., & Pickering, J. P. eds. (1995). Clinical studies of multisensory structured language education. Dallas, TX: *International Multisensory Structured Language Education Council.*

Milde, B. (2014). Unsupervised acquisition of acoustic models for speech-to-text alignment. Master's Thesis, University Technical Darmstat.

Mishra, T., Ljolje, A., & Gilbert, M. (2011). Predicting Human Perceived Accuracy of ASR Systems. In *Annual Conference of the International Speech Communication Association, 12*, 1945-1948.

Mohammad, W., Ruzanna, W. M., Vijayaletchumy, S., Aziz, A., Yasran, A., & Rahim, N. A. (2011). Dyslexia in the aspect of Malay language spelling. *International Journal of Humanities and Social Science (IJHSS), 21*(1), 266-268.

Mostow, J. (2006). Is ASR accurate enough for automated reading tutors, and how can we tell? In *International Conference on Spoken Language Processing. (ICSLP), 9.*

Mporas, I., T. Ganchev, & Fakotakis, N. (2010). Speech segmentation using regression fusion of boundary predictions. *Computer Speech & Language*, *24*(2), 273-288.

Mustafa, M. B., Rosdi, F., Salim, S. S., & Mughal, M. U. (2015). Exploring the Influence of General and Specific Factors on the Recognition Accuracy of an ASR System for Dysarthric Speaker. *Expert Systems with Applications*, *42*, 3924-3932.

Naghibi, T., Hofmann, S., & Pfister, B. (2013). An efficient method to estimate pronunciation from multiple utterances. In *Interspeech Annual Conference of the International Speech Communication Association*, *14*, 1951-1955.

Necibi, K., & Bahi, H. (2012). An Arabic mispronunciation detection system by means of automatic speech recognition technology. In *the International Arab Conference on Information Technology Proceedings*, *13*, 304-308.

Newton, J. M., & Thomas, E. M. (1974). Dyslexia A Guide for Teachers and Parents. *London: University Press.*

Novotney, S., & Callison-Burch, C. (2010). Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *Human Language*

Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 207-215.

Ong, H. F., & Ahmad, A. M. (2011). Malay Language Speech Recognizer with Hybrid Hidden Markov Model and Artificial Neural Network (HMM/ANN). In *International Journal of Information and Education Technology*, *1*(2), 114-119.

Passy, C. (2008). Turning audio into words on the screen. Retrieved January 25, 2015, from http://www.wsj.com/articles/SB122351860225518093.

Pedersen, J. S., & Larsen, L. B. (2010). A Speech Corpus for Dyslexic Reading Training. *Proceedings of the International Conference on Language Resources and Evaluation* (*LREC),* European Language Resources Association, *7,* 2820-2823.

Perea, M., Jimenez, M., Suarez C. P., Fernandez, N., Vina, C., & Cuetos, F. (2014). Ability for voice recognition is a marker for dyslexia in children.

Picone, J., Ganapathiraju, A., & Hamaker, J. (2006). Applications of Kernel Theory to speech. Recognition. *Kernel Methods in Bioengineering, Signal and Image Processing*, 224-240.

Pieraccini, R. (2012). The voice in the machine: Building computers that understand speech Massachusetts Institute of Technology (MIT Press), Cambridge, *141*.

Rabiner, L. R., & Juang, B. H. (1993). Fundamentals of speech recognition, prentice-hall, Englewood.

Radi. M. I. H. (2012). Phonetic transcription: A comparison between manual and automated approach. Master Thesis's, Universiti Utara Malaysia.

Rahman, F. D., Mohamed, N., Mustafa, M. B., & Salim, S. S. (2014). Automatic speech recognition system for Malay speaking children. In *ICT International Student Project Conference (ICT-ISPC)*, *3*, 79-82.

Ramesh, K. V., & Gahankari, S. (2013). Hybrid Artificial Neural Network and Hidden Markov Model (ANN/HMM) for speech and speaker recognition. In *International conference on Green Computing and Technology*, 24-27.

Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models / An Aligner for German. In *Proceedings of ELSNET Goest East and IMACS Workshop*, Moscow, Russia. Retrieved January, 23, 2015, from http://www.ims.uni-stuttgart.˜de/rapp/.

Rasmussen, M. H., Tan, Z. H., Lindberg, B., & Jensen, S. H. (2009). A System for Detecting Miscues in Dyslexic Read Speech. In *Annual Conference of the International Speech Communication Association*, *10*, 1467-1470.

Rello, L., & Llisterri, J. (2012). There are phonetic patterns in vowel substitution errors in texts written by persons with dyslexia. In *Annual World Congress on Learning Disabilities. Learning disabilities: Present and future*, Oviedo, Spain. *21*, 327-38.

Riley, M., Byrne, W., Finke, M., Khudanpur, S., Ljolje, A., McDonough, J., et al. (1999). Stochastic pronunciation modeling from hand-labelled phonetic corpora. *Speech Communication*, *29*(2), 209-224.

Rosdi, F., & Ainon, R. N. (2008). Isolated Malay speech recognition using Hidden Markov Models. *Proceedings of the International Conference on Computer and Communication Engineering*, 721-725.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., et al. (1996). Application of automatic speech recognition to speech and language development in young children. In *Proceedings spoken language of the International Conference on Spoken Language Processing*, Philadelphia, *1*, 176-179.

Saraclar, M., & Khundanpur, S. (2004). Pronunciation change in conversational speech and its implications for automatic speech recognition. In *Computer, Speech and Language, 18*, 375-395.

Sarma, H., Saharia, N., & Sharma, U. (2014). Development of Assamese speech corpus and automatic transcription using HTK. In *Advances in Signal Processing and Intelligent Recognition Systems*. Springer International Publishing, *264*, 119-132.

Sawyer, D. J., Wade, S., & Kim, J. K. (1999). Spelling errors as a window on variations in phonological deficits among students with dyslexia. *Annals of Dyslexia, 49*, 137 - 159.

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*(1), 96-109.

Serridge, B. (2014). An Undergraduate Course on Speech Recognition Based on the CSLU Toolkit. *In International Conference on Spoken Language Processing*, Sydney, Australia, *5*.

Shire, M. L. (2001). Relating frame accuracy with word error in hybrid ANN-HMM ASR. In *Proceedings of the European Conference on Speech Communication and Technology*, *7*, 1797-1800.

Shrawankar, U., & Mahajan, A. (2013). Speech: A Challenge to Digital Signal Processing Technology for Human-to-Computer Interaction. *arXiv preprint arXiv:1305.1925*. 206-212.

Silber, V., & Geri, N. (2014). Can automatic speech recognition be satisfying for audio/video search? Keyword-focused analysis of Hebrew automatic and manual transcription. *Online Journal of Applied Knowledge Management*, *2*(1), 104-121.

Sjolander, K. (2003). An HMM-based system for automatic segmentation and alignment of speech. In *Proceedings of Fonetik*, 93-96.

Sjolander, K., & Beskow, J. (2006). WaveSurfer user manual. Retrieved April 9, 2015, from https://www.speech.kth.se/wavesurfer/man.html.

Sperber, M. (2012). Efficient speech transcription through respeaking. Master's Thesis, Karlsruhe Institute of Technology Department of Computer Science.

Stolcke, A., Ryant, N., Mitra, V., Yuan, J., Wang, W., & Liberman, M. (2014). Highly accurate phonetic segmentation using boundary correction models

and system fusion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 14*, 5552-5556.

Sutton, S., Cole, R. A., De Villiers, J., Schalkwyk, J., Vermeulen, P. J., Macon, M. W., et al. (1998). In *Proceedings of the International Conference on Spoken Language Processing (ICSLP), 98*, 3221-3224.

Taileb, M., Al-Saggaf, R., Al-Ghamdi, A., Al-Zebaidi, M., & Al-Sahafi, S. (2013). YUSR: speech recognition software for dyslexics. *Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience*, Springer Berlin Heidelberg. *8013*, 296-303.

Ting, C. M. (2007). Malay continuous speech recognition using continuous density Hidden Markov Model. Doctoral dissertation, Faculty of Electrical Engineering, Universiti Teknologi Malaysia.

Ting, C. M., & Hussain, S. H., Tan, S. T., & Ariff, A. K. (2007). Automatic phonetic segmentation of Malay speech database. In *International Conference on Information, Communications & Signal Processing, 6*, 1-4.

Tjalve, M., & Huckvale, M. (2005). Pronunciation variation modelling using accent features. In *Proceedings of Euro Speech, Speech Communication, 50*, 605-615.

Togneri, R., Alder, M. D., & Attikiouzel, Y. (1990). Speech processing using artificial neural networks. In *Proceedings of the Australian International Conferences on Speech Science and Technology, 3*, 304-309.

Tolba, M. F., Nazmy, T., Abdelhamid, A. A., & Gadallah, M. E. (2005). A novel method for Arabic consonant/vowel segmentation using wavelet transform.

International Journal on Intelligent Cooperative Information Systems, IJICIS, 5(1), 353-364.

Toth, L., & Kocsor, A. (2007). A segment-based interpretation of HMM/ANN hybrids. *Computer Speech and Language, 21*, 562-578.

Van Bael, C., Boves, L., Heuvel, H. & Strik, H. (2007). Automatic Phonetic Transcription of Large Speech Corpora. *Centre for Language and Speech Technology (CLST)*, Netherlands, *21*(4), 652-668.

Vasilescu, I., Vieru, B., & Lamel, L. (2014). Exploring pronunciation variants for Romanian speech-to-text transcription. In *Spoken Language Technologies for Under-Resourced Languages (SLTU).*St. Petersburg, Russia, 162-168.

Vijayalakshmi, A. (2012). Implementation of Forced Alignment Algorithm For Large Malay Database. Undergraduate Project's Paper, Universiti Teknologi Malaysia.

Wang, Y. Y., Acero, A., & Chelba, C. (2003). Is word error rate a good indicator for spoken language understanding accuracy? In *Automatic Speech Recognition and Understanding (ASRU).* IEEE Workshop, *3*, 577-582.

Wells, J. C. (2006). Phonetic transcription and analysis. *Encyclopaedia of Language and Linguistics.* Amsterdam: Elsevier, 386-396.

Wester, M. (2003). Pronunciation modelling for ASR knowledge based and data derived methods. In *Computer Speech and Language, 17*(1), 69-85.

Williams, J. D., Melamed, I. D., Alonso, T., Hollister, B., & Wilpon, J. (2011). Crowd-sourcing for difficult transcription of speech. In *Automatic Speech Recognition and Understanding* (ASRU), IEEE Workshop. 535-540.

Wise, B., Cole, R., Van V, S., Schwartz, S., Snyder, L., Ngampatipatpong, N., et al., (2005). Learning to read with a virtual tutor: Foundations to literacy. *Interactive literacy education: Facilitating literacy environments through technology*, 31-75.

Wothke, K. (1993). Morphologically based automatic phonetic transcription. *IBM systems Journal*, *32*, 486-511.

Yang, H., Oehlke, C., & Meinel, C. (2011). German speech recognition: A solution for the analysis and processing of lecture recordings. In *International Conference on Computer and Information Science (ICIS)*, *10*, 201-206.

Yoon, S. Y., Chen, L., & Zechner, K. (2010). Predicting word accuracy for the automatic speech recognition of non-native speech. In *Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan, *11*, 773-776.

Yu, K., Gales, M., Wang, L., & Woodland, P. C. (2010). Unsupervised training and directed manual transcription for LVCSR. *Speech Communication*, *52*(7), 652-663.

Yuan, J., & Liberman, M. (2011). Automatic detection of "g-dropping" in American English using forced alignment. In *IEEE Workshop on Automatic Speech Recognition & Understanding*, 490-493.

Yuan, J., Ryant, N., Liberman, M., Stolcke, A., Mitra, V., & Wang, W. (2013). Automatic phonetic segmentation using boundary models. In *Interspeech Annual Conference of the International Speech Communication Association*. 2306-2310.

Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S., & Houtgast, T. (2008). The benefit obtained from visually displayed text from an automatic speech recognizer during listening to speech presented in noise. *Ear and hearing*, *29*(6), 838-852.