MULTINOMIAL LOGISTIC REGRESSION PROBABILITY RATIO-BASED FEATURE VECTORS FOR MALAY VOWEL RECOGNITION

ABDULWAHAB FUNSHO ATANDA

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2021

CERTIFICATION OF THESIS

Permission to Use

In presenting this thesis in fulfillment 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 Vokal adalah sebahagian daripada sistem pengecaman pertuturan automatik (ASR) yang mengelaskan gelombang pertuturan mengikut kumpulan vokal. Prestasi pengecaman vokal Melayu (MVR) seperti masalah pengelasan pelbagai kelas banyak bergantung pada vektor-vektor fitur (FVs). Vektor-vektor fitur (FVs) seperti Pekali Cepstral Mel-Frekuensi (MFCC) telah menghasilkan kadar ralat yang tinggi disebabkan maklumat fonem yang lemah. Fitur kebarangkalian pengelasan yang diubah telah terbukti menjadi alternatif yang lebih baik dalam membawa maklumat fonem. Walau bagaimanapun, dimensi tinggi dalam fitur kebarangkalian memperkenalkan kerumitan tambahan yang menyebabkan kemerosotan prestasi ASR. Tujuan kajian ini adalah untuk meningkatkan prestasi MVR dengan mencadangkan algoritma yang mengubah vektor-vektor fitur MFCC menjadi sekumpulan fitur baharu menggunakan Regresi Logistik Multinomial (MLR) bagi mengurangkan dimensi fitur kebarangkalian. Kajian ini dilakukan dalam empat fasa iaitu pra-pemprosesan dan pengekstrakan fitur, penghasilan pekali regresi terbaik, transformasi fitur, dan penilaian prestasi. Korpus pertuturan terdiri daripada 1953 sampel lima vokal bahasa Melayu iaitu /a/, /e/, /i/, /o/ dan /u/ yang dirakam daripada pelajar dari dua buah universiti awam di Malaysia. Dua set algoritma telah dibangunkan iaitu DBRCs dan FELT. Algoritma DBRCs menentukan pekali regresi terbaik (DBRCs) bagi mendapatkan set pekali regresi (RCs) terbaik daripada vektor-vektor fitur 39-MFCC yang diekstrak melalui pendekatan pensampelan semula dan pertukaran data. Algoritma FELT pula mengubah vektor-vektor fitur 39-MFCC menggunakan kaedah transformasi logistik menjadi vektor-vektor fitur FELT. Kadar pengecaman vokal vektor-vektor fitur FELT dan 39-MFCC dibandingkan dengan menggunakan empat teknik pengelasan yang berbeza iaitu Rangkaian Neural Buatan, MLR, Analisis Diskriminan Linear, dan k-Jiran Terdekat. Hasil pengelasan menunjukkan bahawa vektor-vektor fitur FELT mengatasi prestasi vektor-vektor fitur 39-MFCC dalam MVR. Bergantung pada pengelasan yang digunakan, peningkatan prestasi 1.48% - 11.70% dicapai oleh FELT berbanding MFCC. FELT juga menunjukkan peningkatan ketepatan pengecaman vokal /o/ dan /u/ masing-masing sebanyak 5.13% dan 8.04%. Kajian ini telah menyumbangkan dua algoritma yang dapat menentukan set RC terbaik dan menghasilkan vektor-vektor fitur FELT dari MFCC. Vektor-vektor fitur FELT tidak memerlukan pengurangan dimensi tetapi menghasilkan prestasi yang setanding. Vektor-vektor fitur FELT juga dapat meningkatkan MVR untuk kelima-lima vokal terutamanya /o/ dan /u/. Prestasi MVR yang lebih baik akan mendorong perkembangan sistem berasaskan pertuturan yang menggunakan bahasa Melayu terutamanya untuk masyarakat Malaysia.

Kata Kunci : Pengecaman Pertuturan Automatik, Pekali Regresi, Maklumat Fonem, Transformasi Vektor Fitur, Pengurangan Dimensi.

Abstract

Vowel Recognition is a part of automatic speech recognition (ASR) systems that classifies speech signals into groups of vowels. The performance of Malay vowel recognition (MVR) like any multiclass classification problem depends largely on Feature Vectors (FVs). FVs such as Mel-frequency Cepstral Coefficients (MFCC) have produced high error rates due to poor phoneme information. Classifier transformed probabilistic features have proved a better alternative in conveying phoneme information. However, the high dimensionality of the probabilistic features introduces additional complexity that deteriorates ASR performance. This study aims to improve MVR performance by proposing an algorithm that transforms MFCC FVs into a new set of features using Multinomial Logistic Regression (MLR) to reduce the dimensionality of the probabilistic features. This study was carried out in four phases which are pre-processing and feature extraction, best regression coefficients generation, feature transformation, and performance evaluation. The speech corpus consists of 1953 samples of five Malay vowels of /a/, /e/, /i/, /o/ and /u/ recorded from students of two public universities in Malaysia. Two sets of algorithms were developed which are DBRCs and FELT. DBRCs algorithm determines the best regression coefficients (DBRCs) to obtain the best set of regression coefficients (RCs) from the extracted 39-MFCC FVs through resampling and data swapping approach. FELT algorithm transforms 39-MFCC FVs using logistic transformation method into FELT FVs. Vowel recognition rates of FELT and 39-MFCC FVs were compared using four different classification techniques of Artificial Neural Network, MLR, Linear Discriminant Analysis, and k-Nearest Neighbour. Classification results showed that FELT FVs surpass the performance of 39-MFCC FVs in MVR. Depending on the classifiers used, the improved performance of 1.48% - 11.70% was attained by FELT over MFCC. Furthermore, FELT significantly improved the recognition accuracy of vowels /o/ and /u/ by 5.13% and 8.04% respectively. This study contributes two algorithms for determining the best set of RCs and generating FELT FVs from MFCC. The FELT FVs eliminate the need for dimensionality reduction with comparable performances. Furthermore, FELT FVs improved MVR for all the five vowels especially /o/ and /u/. The improved MVR performance will spur the development of Malay speech-based systems, especially for the Malaysian community.

Keywords: Automatic Speech Recognition, Regression coefficients, Phoneme Information, Feature Vector Transformation, Dimensionality Reduction.

Acknowledgment

First and foremost, my gratitude goes to the Almighty Allah The Most Merciful, The Most Benevolent for HIS unending mercies. And to my beloved mother who stops living due to my absence while pursuing this Ph.D. May Allah reward and grants her long life.

My profound gratitude goes to my supervisor Assoc. Prof. Dr. Shahrul Azmi Mohd Yusof and Assoc. Prof. Dr. Husniza Husni for their immense guidance, tutorage, patience, encouragement, and generosity that facilitated the completion of this thesis. Deserving of gratitude is Assoc. Prof. Dr. Nor Idayu Mahat for her immeasurable contributions towards the success of this thesis. Also worthy of my gratitude are the panelists of my VIVA session – Prof. Dr. Suhaidi Bin Hassan, Prof. Dr. Zainab Abu Bakar, and Assoc. Prof. Dr. Siti Sakira for their critical and objective contributions that shaped the final thesis. To the management and the entire staff of both the School of Computing and Awang Had Salleh Graduate School, I sincerely appreciate your contributions. My gratitude to the following individuals: an unknown UAE citizen, Al-Homery, and Siti Safinas for their generosity that sees to the smooth completion of this thesis. I would like to express my appreciation and gratitude to everyone who has contributed to completing this thesis, especially Asiribo, Nurudeen, Ahmad, Abdulkareem for their encouragement, prayers, and constant visit to my family back in Nigeria. And to all my colleagues in IASDO and all those I met during my course of study in UMM and stay in Malaysia such as Muen,Naseer, Shehu, Sakira, and Hapini, I say a big thank you to all.

I would like also to thank my parents, siblings, and all of my relatives for their love and support. Last but not the least, I am very grateful and indebted to my loving wife Safurat and daughters Shekinah, Abeebah, Abeedah, and Ameedah for their patience, endurance, prayers, and encouragement while undertaking this study. To them, I say a very big thank you and I love you all.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

ASR	Automatic Speech Recognition
BI	Bark Intensity
CM	Confusion Matrix
CRs	Classification Rates
DBRCs	Determine Best Regression Coefficients
ER	Error Rate
FE	Feature Extraction
FELT	Feature Extraction by Logistic Regression
FVs	Feature Vectors
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HSR	Human speech recognition
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LR	Linear Regression
LRL	Low Resource Language
MFCC	Mel-frequency Cepstral Coefficient
MLR	Multinomial Logistic Regression
MVR	Malay Vowel Recognition
PCA	Principal Component Analysis
PLP	Perceptual Linear Prediction
RCs	Regression Coefficients
VR	Vowel Recognition

List of Appendices

CHAPTER ONE
INTRODUCTION

1.0 Background of the Study

Multinomial Logistic Regression (MLR) is a discriminative and probabilistic classification model that belongs to a larger class of logistic regression (LR) models (Hansen & Liu, 2016; Kayabol, 2019). It predicts the occurrence of an event given some sets of training data by computing the probability of the likely occurrence of events. In making the prediction outcomes for a multiclass event, MLR uses several predictor variables that can either be numerical or categorical (Abdalmalak & Gallardo-Antolín, 2016; Gunduz & Karacan, 2017). The output of MLR is usually computed as the probability of success divided by the probability of failure. The outcome of the regression analysis is in the form of probability ratios for each of the events.

In making the final prediction decision, the event that has the highest probability ratio is predicted as a successful outcome. Although, the event with the highest probability ratio is predicted, nonetheless, probability values are also assigned to the remaining unpredicted events. The sum of the probability ratios for all the possible outcomes adds up to a total sum of 1.0, with the values of each of the events ranging between 0 and 1.0 (Elfadaly & Garthwaite, 2020). The probability ratios are the ratios between the likelihood numbers of events outcomes. The probability ratios form the basis for prediction decisions, and at the same time can be used for the derivation of the odds ratio that forms the basis for interpreting the output of MLR (Finch, Bolin, & Kelley, 2019; Weisburd & Britt, 2014). MLR is mostly being used as a classifier in vowel

REFERENCES

Abad, A., Bell, P., Carmantini, A., & Renais, S. (2020). *Cross lingual transfer learning for zero-resource domain adaptation.* Paper presented at the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Abdalmalak, K. A., & Gallardo-Antolín, A. (2016). Enhancement of a text-independent speaker verification system by using feature combination and parallel structure classifiers. *Neural Computing and Applications*. doi: 10.1007/s00521-016-2470-x

Ahmed, A. I., Chiverton, J. P., Ndzi, D. L., & Becerra, V. M. (2019). Speaker recognition using PCA-based feature transformation. *Speech Communication, 110*, 33-46.

Ahmed, H., & Nandi, A. K. (2018). Compressive sampling and feature ranking framework for bearing fault classification with vibration signals. *IEEE Access, 6*, 44731-44746.

Àkànbí, L. A., & Odéjobí, O. À. (2011). Automatic recognition of oral vowels in tone language: Experiments with fuzzy logic and neural network models. *Applied Soft Computing, 11*, 1467-1480.

Alasadi, A. A., Aldhayni, T., Deshmukh, R. R., Alahmadi, A. H., & Alshebami, A. S. (2020). Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System. *Eng. Technol. Appl. Sci. Res, 10*(2), 5547-5553.

Alharbi, S. (2018). *Automatic Framework to Aid Therapists to Diagnose Children who Stutter.* University of Sheffield.

Ali, Z., Muhammad, G., & Alhamid, M. F. (2017). An automatic health monitoring system for patients suffering from voice complications in smart cities. *IEEE Access, 5*, 3900-3908.

Alías, F., Socoró, J., & Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences, 6*(5), 143.

Alim, S. A., & Rashid, N. K. A. (2018). Some commonly used speech feature extraction algorithms. *From Natural to Artificial Intelligence-Algorithms and Applications*.

Almisreb, A. A. (2016). Consonants recognition and noise reduction for Arabic phonemes based Malay speakers.

Aloqayli, F. M., & Alotaibi, Y. A. (2017). *Spoken Arabic Vowel Recognition Using ANN.* Paper presented at the 2017 European Modelling Symposium (EMS).

Alqudah, A. M., Albadarneh, A., Abu-Qasmieh, I., & Alquran, H. (2019). Developing of robust and high accurate ECG beat classification by combining Gaussian mixtures and wavelets features. *Australasian physical & engineering sciences in medicine*, *42*(1), 149-157.

Altay, O., & Ulas, M. (2018). *Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children.* Paper presented at the 2018 6th International Symposium on Digital Forensic and Security (ISDFS).

Amami, R., Ayed, D. B., & Ellouze, N. (2015). An empirical comparison of SVM and some supervised learning algorithms for vowel recognition. *arXiv preprint arXiv:1507.06021*.

Amuda, S., Boril, H., Sangwan, A., & Hansen, J. H. (2010). *Limited resource speech recognition for Nigerian English.* Paper presented at the Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on.

Amuda, S., Boril, H., Sangwan, A., Ibiyemi, T. S., & Hansen, J. H. (2014). Engineering Analysis and Recognition of Nigerian English: An Insight into Low Resource Languages. *Transactions on Machine Learning and Artificial Intelligence*, *2*(4), 115-128.

Anusuya, M., & Katti, S. (2010). Speech recognition by machine, A review. *arXiv preprint arXiv:1001.2267*.

Anusuya, M., & Katti, S. (2011). Classification Techniques used in Speech Recognition Applications: A Review. *Int. J. Comp. Tech. Appl*, *2*(4), 910-954.

Arifin, N., & Tiun, S. (2013). Predicting malay prominent syllable using support vector machine. *Procedia Technology*, *11*, 861-869.

Arora, N. (2016). Automatic Speech Recognition System: A Review. *International Journal of Computer Applications*, *151*(1).

Arslan, L. M., & Hansen, J. H. (1996). Language accent classification in American English. *Speech Communication, 18*(4), 353-367. doi: http://dx.doi.org/10.1016/0167-6393(96)00024-6

Arslan, L. M., & Hansen, J. H. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *The Journal of the Acoustical Society of America*, *102*(1), 28-40.

Atanda, A. F., Azmi, M. Y. S., & Husni, H. (2017). *Acoustic Comparison of Malaysian and Nigerian English Accents*. Paper presented at the The 10th Malaysian Software Engineering Conference (MySEC2017), Universiti Malaysia Terengganu, Malaysia.

Azmi, M. Y. S. (2010a). *Feature Extraction and Classification of Malay Speech Vowels.* Universiti Malaysia Perlis.

Azmi, M. Y. S. (2010b). *Feature extraction and classification of malay speech vowels.* School of Mechatronic Engineering, Universiti Malaysia Perlis.

Azmi, M. Y. S. (2014). An improved feature extraction method for Malay vowel recognition based on spectrum delta. *International Journal of Software Engineering and Its Applications, 8*(1), 413-426.

Azmi, M. Y. S. (2016a). Development of Malay Word Pronunciation Application using Vowel Recognition. *Malay, 9*(1).

Azmi, M. Y. S. (2016b). Malay Word Pronunciation Test Application for Pre-School Children. *Int Journal Of Interactive Digital Media, 4*(2), 2289-4098.

Azmi, M. Y. S., Atanda, A. F., & Hariharan, M. (2013). *A review of Yorùbá Automatic Speech Recognition.* Paper presented at the System Engineering and Technology (ICSET), 2013 IEEE 3rd International Conference on.

Azmi, M. Y. S., Idayu, M. N., Roshidi, D., Yaakob, A., & Yaacob, S. (2012). Noise Robustness of Spectrum Delta (SpD) Features in Malay Vowel Recognition *Computer Applications for Communication, Networking, and Digital Contents* (pp. 270-277): Springer.

Azmi, M. Y. S., Siraj, F., Yaacob, S., Paulraj, M., & Nazri, A. (2010a). *Improved Malay Vowel Feature Extraction Method Based on First and Second Formants*. Paper presented at the Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on.

Azmi, M. Y. S., Siraj, F., Yaacob, S., Paulraj, M., & Nazri, A. (2010b). *Improved Malay Vowel Feature Extraction Method Based on First and Second Formants.* Paper presented at the 2010 Second International Conference on Computational Intelligence, Modelling and Simulation.

Azmi, M. Y. S., & Yaacob, S. (2008). Classification of Malaysian vowels using formant based features. *Journal of ICT, 7,* 27-40.

Azmi, M. Y. S., Yaacob, S., & Paulraj, M. (2009). Vowel recognition using First Formant Feature.

Azmi, M. Y. S., Yaacob, S., Paulraj, M. P., & Nazri, A. (2010). *Improved Malay vowel feature extraction method based on first and second formants*. Paper presented at the Second International Conference on Computational Intelligence, Modelling and Simulation.

Baart, A., Bon, A., de Boer, V., Tuyp, W., & Akkermans, H. (2018). *Ney Yibeogo-Hello World: A Voice Service Development Platform to Bridge the Web's Digital Divide.* Paper presented at the WEBIST.

Balaji, V., & Sadashivappa, G. (2019). *Waveform Analysis and Feature Extraction from Speech Data of Dysarthric Persons.* Paper presented at the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN).

Batliner, A., & Möbius, B. (2019). Prosody in automatic speech processing: The Oxford Handbook of Prosody, C. Gussenhoven and A. Chen, Eds. Oxford, UK ….

Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication, 56,* 85-100.

Best, H., & Wolf, C. (2015). Logistic regression *The SAGE handbook of regression analysis and causal inference* (pp. 153-171): Sage Los Angeles, CA.

Bhatt, S., Dev, A., & Jain, A. (2018). *Hindi Speech Vowel Recognition Using Hidden Markov Model*. Paper presented at the SLTU.

Bhatt, S., Dev, A., & Jain, A. (2020). Confusion analysis in phoneme based speech recognition in Hindi. *Journal of Ambient Intelligence and Humanized Computing*, 1-26.

Bhowmik, T., & Mandal, S. K. D. (2018). Segmental and Supra-Segmental Feature Based Speech Recognition System forUnder Resourced Languages.

Biem, A., Katagiri, S., & Juang, B.-H. (1997). Pattern recognition using discriminative feature extraction. *IEEE Transactions on Signal Processing, 45*(2), 500-504.

Birkenes, Ø. (2007). A framework for speech recognition using logistic regression.

Bishop, C. M. (2006). *Pattern recognition and machine learning*: springer.

Bonev, B., Escolano, F., & Cazorla, M. (2008). Feature selection, mutual information, and the classification of high-dimensional patterns. *Pattern Analysis and Applications, 11*(3-4), 309-319.

Bora, A., Vajpai, J., & Gaur, S. B. (2019). Phonology based Fuzzy Phoneme Recognition. *Available at SSRN 3358093*.

Bourouba, H., & Djemili, R. (2020). FEATURE EXTRACTION ALGORITHM USING NEW CEPSTRAL TECHNIQUES FOR ROBUST SPEECH RECOGNITION. *Malaysian Journal of Computer Science, 33*(2), 90-101.

Campbell, E., Phinyomark, A., & Scheme, E. (2019). *Linear discriminant analysis with bayesian risk parameters for myoelectric control*. Paper presented at the 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP).

Cardinal, P., Dehak, N., Zhang, Y., & Glass, J. (2015). *Speaker adaptation using the i-vector technique for bottleneck features*. Paper presented at the Sixteenth Annual Conference of the International Speech Communication Association.

Chamroukhi, F., & Huynh, B. T. (2018). *Regularized maximum-likelihood estimation of mixture-of-experts for regression and clustering.* Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN).

Chan, K. Y., & Hall, M. D. (2019). The importance of vowel formant frequencies and proximity in vowel space to the perception of foreign accent. *Journal of Phonetics, 77*, 100919.

Chandra, M. (2017). *DSCC features for Hindi vowel classification.* Paper presented at the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET).

Chang, Y.-p., Chang, R. Y., Lin, C.-Y., & Luo, X. (2016). Mandarin tone and vowel recognition in cochlear implant users: Effects of talker variability and bimodal hearing. *Ear and hearing, 37*(3), 271.

Chen, W., & Zhang, X. (2017). *Research on text categorization model based on LDA—KNN.* Paper presented at the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).

Chittaragi, N. B., & Koolagudi, S. G. (2019). Acoustic-phonetic feature based Kannada dialect identification from vowel sounds. *International Journal of Speech Technology, 22*(4), 1099-1113.

Chiu, C.-C., Tripathi, A., Chou, K., Co, C., Jaitly, N., Jaunzeikare, D., . . . Sankar, A. (2017). Speech recognition for medical conversations. *arXiv preprint arXiv:1711.07274*.

Cincarek, T. (2008). Selective training for cost-effective development of real-environment speech recognition applications. *PhD thesis, Nara Institute of Science and Technology*.

Collins, J. T. (1989). Malay dialect research in Malaysia: the issue of perspective. *Bijdragen tot de Taal-, Land-en Volkenkunde*(2/3de Afl), 235-264.

Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems with Applications, 39*(8), 7562-7573.

Dalmia, S., Li, X., Metze, F., & Black, A. W. (2018). *Domain robust feature extraction for rapid low resource asr development.* Paper presented at the 2018 IEEE Spoken Language Technology Workshop (SLT).

Dalmia, S., Sanabria, R., Metze, F., & Black, A. W. (2018). *Sequence-based multi-lingual low resource speech recognition.* Paper presented at the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Daniel, J., & Martin, J. (2009). Speech and Language Processing-Edition: 2. *Prentice-Hall Inc, ISBN, 131873210*, 2-16,489.

Daoud, J. I. (2017). *Multicollinearity and regression analysis.* Paper presented at the Journal of Physics: Conference Series.

Darlington, R. B., & Hayes, A. F. (2016). *Regression analysis and linear models: Concepts, applications, and implementation*: Guilford Publications.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on, 28*(4), 357-366.

de-La-Calle-Silos, F., & Stern, R. M. (2017). Synchrony-based feature extraction for robust automatic speech recognition. *IEEE Signal Processing Letters, 24*(8), 1158-1162.

de Jong, V. M., Eijkemans, M. J., van Calster, B., Timmerman, D., Moons, K. G., Steyerberg, E. W., & van Smeden, M. (2019). Sample size considerations and predictive performance of multinomial logistic prediction models. *Statistics in medicine, 38*(9), 1601-1619.

Deller, J. R., Proakis, J. G., & Hansen, J. H. (2000). *Discrete-time processing of speech signals*: Ieee New York, NY, USA:.

DeMarco, A., & Cox, S. J. (2013). *Native accent classification via i-vectors and speaker compensation fusion.* Paper presented at the INTERSPEECH.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing, 7*(3–4), 197-387.

Desai, V. A., & Thakar, V. K. (2017). Algorithm for Keyword Spotting with Application to Speech Recognition.

Deshpande, S., Chikkerur, S., & Govindaraju, V. (2005). *Accent classification in speech.* Paper presented at the Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on.

Devi, K. J., & Thongam, K. (2019). Automatic speaker recognition with enhanced swallow swarm optimization and ensemble classification model from speech signals. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.

Devroye, L., Györfi, L., & Lugosi, G. (2013). *A probabilistic theory of pattern recognition* (Vol. 31): Springer Science & Business Media.

Dey, S., & Alam, M. A. (2018). *Formant Based Bangla Vowel Perceptual Space Classification Using Support Vector Machine and K-Nearest Neighbor Method.* Paper presented at the 2018 21st International Conference of Computer and Information Technology (ICCIT).

Disatnik, D., & Sivan, L. (2016). The multicollinearity illusion in moderated regression analysis. *Marketing Letters, 27*(2), 403-408.

Djamal, E. C., Nurhamidah, N., & Ilyas, R. (2017). *Spoken word recognition using mfcc and learning vector quantization.* Paper presented at the 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI).

Dodia, S., Edla, D. R., Bablani, A., Ramesh, D., & Kuppili, V. (2019). An efficient EEG based deceit identification test using wavelet packet transform and linear discriminant analysis. *Journal of neuroscience methods, 314*, 31-40.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification. *New York: John Wiley, Section, 10,* l.

Edwards, E., Salloum, W., Finley, G. P., Fone, J., Cardiff, G., Miller, M., & Suendermann-Oeft, D. (2017). *Medical speech recognition: reaching parity with humans.* Paper presented at the International Conference on Speech and Computer.

Elfadaly, F. G., & Garthwaite, P. H. (2020). On quantifying expert opinion about multinomial models that contain covariates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Fagerland, M. W., & Hosmer, D. W. (2017). How to test for goodness of fit in ordinal logistic regression models. *The Stata Journal, 17*(3), 668-686.

Fang, Q. (2019). *Is average RMSE appropriate for evaluating acoustic-to-articulatory inversion?* Paper presented at the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC).

Fantaye, T. G., Yu, J., & Hailu, T. T. (2019). *Syllable-based Speech Recognition for a Very Low-Resource Language, Chaha.* Paper presented at the Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence.

Feng, J., Ramabhadran, B., Hansen, J. H. L., & Williams, J. D. (2012). Trends in Speech and Language Processing [In the Spotlight]. *Signal Processing Magazine, IEEE, 29*(1), 177-179. doi: 10.1109/msp.2011.943131

Ferreira, E., Nocera, P., Goudi, M., & Thi, N. D. D. (2012). *YAST: A scalable ASR toolkit especially designed for under-resourced languages.* Paper presented at the Asian Language Processing (IALP), 2012 International Conference on.

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC medical informatics and decision making, 12*(1), 8.

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R*: Crc Press.

Flatt, C., & Jacobs, R. L. (2019). Principle Assumptions of Regression Analysis: Testing, Techniques, and Statistical Reporting of Imperfect Data Sets. *Advances in Developing Human Resources, 21*(4), 484-502.

Foley, B., Arnold, J. T., Coto-Solano, R., Durantin, G., Ellison, T. M., van Esch, D., . . . Nash, D. (2018). *Building Speech Recognition Systems for Language Documentation: The CoEDL Endangered Language*

Pipeline and Inference System (ELPIS). Paper presented at the SLTU.

Fook, C., Hariharan, M., Yaacob, S., & Adom, A. (2012). *A review: Malay speech recognition and audio visual speech recognition.* Paper presented at the Biomedical Engineering (ICoBE), 2012 International Conference on.

Fook, C., Hariharan, M., Yaacob, S., & Ah, A. (2012). *Malay speech recognition in normal and noise condition.* Paper presented at the Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on.

Forgie, J. W., & Forgie, C. D. (1959). Results obtained from a vowel recognition computer program. *The Journal of the Acoustical Society of America*, *31*(11), 1480-1489.

Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017). Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks. *The Journal of the Acoustical Society of America, 142*(2), 1025-1033.

Fry, D. B., & Denes, P. (1959). The Design and Operation of the Mechanical Speech Recognizer at University College London. *J. British Inst. Radio Engr*, *19*(4), 211-229.

Fu, K.-S. (2019). *Applications of pattern recognition*: CRC press.

Gales, M., & Young, S. (2008). The application of hidden Markov models in speech recognition. *Foundations and Trends® in Signal Processing, 1*(3), 195-304.

Garcia, L. P., Lorena, A. C., de Souto, M. C., & Ho, T. K. (2018). *Classifier recommendation using data complexity measures.* Paper presented at the 2018 24th International Conference on Pattern Recognition (ICPR).

Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., & Schwenk, H. (2005). *Where are we in transcribing French broadcast news?* Paper presented at the INTERSPEECH.

Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics, 49*(3), 291-304.

Ghesquiere, P.-J., & Van Compernolle, D. (2002). *Flemish accent identification based on formant and duration features.* Paper presented at the Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on.

Ghorbani, S., & Hansen, J. H. (2019). Leveraging native language information for improved accented speech recognition. *arXiv preprint arXiv:1904.09038*.

Ghori, K. M., Imran, M., Nawaz, A., Abbasi, R. A., Ullah, A., & Szathmary, L. (2020). Performance analysis of machine learning classifiers for non-technical loss detection. *Journal of Ambient Intelligence and Humanized Computing*, 1-16.

Ghorshi, S., Vaseghi, S., & Yan, Q. (2008). Cross-entropic comparison of formants of British, Australian and American English accents. *Speech Communication, 50*(7), 564-579.

Ghosh, M., & Sanyal, G. (2018). Performance assessment of multiple classifiers based on ensemble feature selection scheme for sentiment analysis. *Applied Computational Intelligence and Soft Computing, 2018*.

Ghosh, T., Saha, S., & Ferdous, A. I. (2016). *Formant Analysis of Bangla Vowel for Automatic Speech Recognition.* Paper presented at the Prec. Signal & Image Processing: An Int. Journal (SIPIJ).

Gonçalves, D. A., Baranauskas, M. C. C., & dos Reis, J. C. (2020). *Accessibility in Pervasive Systems: An Exploratory Study.* Paper presented at the International Conference on Human-Computer Interaction.

Gowdy, J., & Tufekci, Z. (2000). *Mel-scaled discrete wavelet coefficients for speech recognition.* Paper presented at the Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on.

Gray, S. S. (2006). *Speech Science Modeling for Automatic Accent and Dialect Classification.* University of Colorado.

Grézl, F., & Karafiát, M. (2014). *Adapting multilingual neural network hierarchy to a new language*. Paper presented at the Spoken Language Technologies for Under-Resourced Languages.

Guglani, J., & Mishra, A. N. (2018). Continuous Punjabi speech recognition model based on Kaldi ASR toolkit. *International Journal of Speech Technology*, *21*(2), 211-216.

Gunduz, M., & Karacan, H. V. (2017). Assessment of abnormally low tenders: a multinomial logistic regression approach. *Technological and Economic Development of Economy, 23*(6), 848-859.

Gupta, D., Bansal, P., & Choudhary, K. (2018). The state of the art of feature extraction techniques in speech recognition *Speech and language processing for human-machine communications* (pp. 195-207): Springer.

Gupta, H., & Gupta, D. (2016). *LPC and LPCC method of feature extraction in Speech Recognition System*. Paper presented at the 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence).

Gupta, K., & Gupta, D. (2016). *An analysis on LPC, RASTA and MFCC techniques in Automatic Speech recognition system*. Paper presented at the 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence).

Habib, M. R., Hassan, M. A., Seoud, R. A. A., & Bayoumi, A. M. (2017). Mechanical fault detection and classification using pattern recognition based on bispectrum algorithm *Advanced Technologies for Sustainable Systems* (pp. 147-165): Springer.

Hai, V., Xiao, X., Chng, E. S., & Li, H. (2014). Cross-lingual phone mapping for large vocabulary speech recognition of under-resourced languages. *IEICE TRANSACTIONS on Information and Systems, 97*(2), 285-295.

Hamad, S. S., & Alzubaidi, N. M. (2020). African-American Vernacular: Cultural and Literary Identity of Selected African-American Poets. *Journal of Talent Development and Excellence, 12*(2s), 2490-2500.

Hamdy, A., Hefny, H., Salama, M. A., Hassanien, A. E., & Kim, T.-h. (2012). *The importance of handling multivariate attributes in the* *identification of heart valve diseases using heart signals.* Paper presented at the Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on.

Hamid, H. A., Raudhah Ismail, S., Azwal Sulaiman, S., & Amin, N. A. M. (2020). Assessing the Effect of Different Covariates Distributions on Parameter Estimates for Multinomial Logistic Regression (MLR). *MS&E, 767*(1), 012014.

Hanani, A., Russell, M., & Carey, M. J. (2011). *Speech-based identification of social groups in a single accent of British English by humans and computers.* Paper presented at the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4876-4879).

Hanani, A., Russell, M. J., & Carey, M. J. (2013). Human and computer recognition of regional accents and ethnic groups from British English speech. *Computer Speech & Language, 27*, 59-74.

Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems, 140*, 103-119.

Hanifa, R., Isa, K., & Mohamad, S. (2017). *Malay speech recognition for different ethnic speakers: an exploratory study.* Paper presented at the 2017 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE).

Hansen, J. H. (2015). Robust Speech Processing & Recognition: Speaker ID, Language ID, Speech Recognition/Keyword Spotting, Diarization/Co-Channel/Environmental Characterization, Speaker State Assessment: DTIC Document.

Hansen, J. H., & Liu, G. (2016). Unsupervised accent classification for deep data fusion of accent and language information. *Speech Communication, 78*, 19-33.

Haque, M. N. (2017). *Genetic algorithm-based ensemble methods for large-scale biological data classification.* The University of Newcastle.

Haridas, A. V., Marimuthu, R., & Sivakumar, V. G. (2018). A critical review and analysis on techniques of speech recognition: The road ahead. *International Journal of Knowledge-based and Intelligent Engineering Systems, 22*(1), 39-57.

Haug, J. M. M. (2017). *Voice Conversion using Deep Learning.* NTNU.

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3): Prentice Hall New York.

Herff, C., & Schultz, T. (2016). Automatic speech recognition from neural signals: a focused review. *Frontiers in neuroscience, 10*, 429.

Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. *The Journal of the Acoustical Society of America, 87*(4), 1738-1752.

Hermansky, H., Ellis, D. P., & Sharma, S. (2000). *Tandem connectionist feature extraction for conventional HMM systems.* Paper presented at the icassp.

Hevner, A., & Chatterjee, S. (2010). Design research in information systems integrated: Springer.

Hickey, G. L., Kontopantelis, E., Takkenberg, J. J., & Beyersdorf, F. (2019). Statistical primer: checking model assumptions with regression diagnostics. *Interactive cardiovascular and thoracic surgery, 28*(1), 1-8.

Hidayat, R., Yulian, D., Bejo, A., & Sumaryono, S. (2017). *Comparison of vowel feature extraction on time and frequency domain.* Paper presented at the 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE).

Hidayat, S., PN, H. R., & Kumoro, D. T. (2017). Determination of the Optimum Wavelet Basis Function for Indonesian Vowel Voice Recognition. *Jurnal Elektronika dan Telekomunikasi, 17*(2), 42-47.

Hill, D. R., Taube-Schock, C. R., & Manzara, L. (2017). Low-level articulatory synthesis: A working text-to-speech solution and a linguistic tool. *The Canadian Journal of Linguistics/La revue canadienne de linguistique, 62*(3), 371-410.

Hillel, T., Bierlaire, M., Elshafie, M., & Ying, J. (2018). *Validation of probabilistic classifiers*. Paper presented at the 18th Swiss Transport Research Conference (STRC).

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized/hVd/utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*(6), 3509-3523.

Holambe, R. S., & Deshpande, M. S. (2012). Noise robust speaker identification: using nonlinear modeling techniques *Forensic Speaker Recognition* (pp. 153-182): Springer.

Holambe, R. S., & ME, M. S. D. (2012). Noise Robust Speaker Identification: Using Nonlinear Modeling Techniques *Forensic Speaker Recognition* (pp. 153-182): Springer.

Hoshen, Y., Weiss, R. J., & Wilson, K. W. (2015). *Speech acoustic modeling from raw multichannel waveforms.* Paper presented at the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Hosmer, D., & Lemeshow, S. (2004). *Applied logistic regression*. New York: Wiley-Interscience.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.

Hsiao, R., Fuhs, M. C., Tam, Y.-C., Jin, Q., & Schultz, T. (2008). *The CMU-interACT 2008 Mandarin transcription system.* Paper presented at the INTERSPEECH.

Hu, H. (2010). *Nonlinear discriminant analysis based feature dimensionality reduction for automatic speech recognition*: State University of New York at Binghamton.

Huang, X., Acero, A., Hon, H.-W., & Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*: Prentice Hall PTR.

Huckvale, M. (2007). ACCDIST: an accent similarity metric for accent recognition and diagnosis *Speaker Classification II* (pp. 258-275): Springer.

Hung, J.-w., Lin, J.-S., & Wu, P.-J. (2018). Employing robust principal component analysis for noise-robust speech feature extraction in automatic speech recognition with the structure of a deep neural network. *Applied System Innovation, 1*(3), 28.

Inc, S. (2003). *SPSS regression models 12.0*: SPSS Inc.

Jain, A. K., Duin, R. P., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(1), 4-37.

Jalalvand, S., Negri, M., Turchi, M., de Souza, J. G., Falavigna, D., & Qwaider, M. R. (2016). *Transcrater: a tool for automatic speech recognition quality estimation.* Paper presented at the Proceedings of ACL-2016 System Demonstrations.

Jamil, N., Ramli, I., & Ardi, N. (2019). Formant Characteristics of Malay Vowels of Perlis, Kelantan and Terengganu. *Journal of Information & Communication Technology, 18*(4).

Jelinek, F. (2018). Audio Features for Automatic Speech Recognition and Audio Analysis. *Audio and Speech Processing with MATLAB*, 14.

Johansen, F. T., & Johnsen, M. H. (1994). *Non-linear input transformations for discriminative HMMs.* Paper presented at the Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing.

Jorgensen, B. (2019). *Theory of Linear Models*: Routledge.

Juan, S. S., Besacier, L., Lecouteux, B., & Tan, T.-P. (2015). Merging of Native and Non-native Speech for Low-resource Accented ASR *Statistical Language and Speech Processing* (pp. 255-266): Springer.

Kadyan, V., Mantri, A., Aggarwal, R., & Singh, A. (2019). A comparative study of deep neural network based Punjabi-ASR system. *International Journal of Speech Technology*, *22*(1), 111-119.

Kamble, B. C. (2016). Speech Recognition Using Artificial Neural Network–A Review. *Int. J. Comput. Commun. Instrum. Eng, 3*(1), 61-64.

Kanda, N., Horiguchi, S., Fujita, Y., Xue, Y., Nagamatsu, K., & Watanabe, S. (2019). *Simultaneous speech recognition and speaker diarization for monaural dialogue recordings with target-speaker acoustic models*. Paper presented at the 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).

Karim, S., Mohd, F., Musa, H. b. H., Mahmood, A. H., Aman, M. S., Yunus, A. G., & Onn, F. M. (2016). *Tatabahasa dewan*: Dewan Bahasa dan Pustaka.

Karpagavalli, S., & Chandra, E. (2015). *Phoneme and word based model for tamil speech recognition using GMM-HMM*. Paper presented at the Advanced Computing and Communication Systems, 2015 International Conference on.

Karpagavalli, S., & Chandra, E. (2016). A review on automatic speech recognition architecture and approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 9*(4), 393-404.

Kassim, F. N. C., Vijean, V., Muthusamy, H., Abdullah, R., & Abdullah, Z. (2019). *Voice Pathology Analysis using DT-CWPT and ReliefF Algorithm*. Paper presented at the Journal of Physics: Conference Series.

Kat, L. W., & Fung, P. (1999). *Fast accent identification and accented speech recognition*. Paper presented at the Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on.

Kayabol, K. (2019). Approximate Sparse Multinomial Logistic Regression for Classification. *IEEE Transactions on pattern analysis and machine intelligence*.

Keshet, J., & Bengio, S. (2009). *Automatic speech and speaker recognition: Large margin and kernel methods*: John Wiley & Sons.

Kham, N. (2016). *Adaptive Boosting for Automatic Speech Recognition*: Northeastern University.

Khaw, Y. M., & Tan, T. P. (2012). *Pronunciation Modeling for Malaysian English.* Paper presented at the Asian Language Processing (IALP), 2012 International Conference on.

Khwaja, M. K., Vikash, P., Arulmozhivarman, P., & Lui, S. (2016). Robust phoneme classification for automatic speech recognition using hybrid features and an amalgamated learning model. *International Journal of Speech Technology, 19*(4), 895-905.

Kim, M., Oh, Y. R., & Kim, H. K. (2010). Non-native Pronunciation Variation Modeling for Automatic Speech Recognition. *SCIYO. COM,* 83.

Kirby, B. J., Kopun, J. G., Spratford, M., Mollak, C. M., Brennan, M. A., & McCreery, R. W. (2017). Listener performance with a novel hearing aid frequency lowering technique. *Journal of the American Academy of Audiology, 28*(9), 810-822.

Kleinschmidt, T. F. (2010). Robust speech recognition using speech enhancement.

Konig, Y., Heck, L., Weintraub, M., & Sonmez, K. (1998). *Nonlinear discriminant feature extraction for robust text-independent speaker recognition.* Paper presented at the Proc. RLA2C, ESCA workshop on Speaker Recognition and its Commercial and Forensic Applications.

Krug, A., & Stober, S. (2018). *Introspection for convolutional automatic speech recognition.* Paper presented at the Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP.

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access, 5,* 7776-7797.

Lal, P., & King, S. (2013). Cross-lingual automatic speech recognition using tandem features. *IEEE transactions on audio, speech, and language processing, 21*(12), 2506-2515.

Larraza, S., Samuel, A. G., & Oñederra, M. L. (2016). Listening to accented speech in a second language: First language and age of acquisition effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(11), 1774.

Lee, L. W., Low, H. M., & Mohamed, A. R. (2013). A comparative analysis of word structures in Malay and English children's stories. *Social Sciences & Humanities, 21*(1), 67-84.

Lestari, D. P., & Furui, S. (2010). Adaptation to pronunciation variations in Indonesian spoken query-based information retrieval. *IEICE TRANSACTIONS on Information and Systems, 93*, 2388-2396.

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and information systems, 10*(4), 453-472.

Liu, B., Zhang, W., Xu, X., & Chen, D. (2019). *Time Delay Recurrent Neural Network for Speech Recognition.* Paper presented at the Journal of Physics: Conference Series.

Liu, W. K., & Fung, P. N. (2000). *MLLR-based accent model adaptation without accented data.* Paper presented at the Sixth International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, 16-20 October 2000, vol. 3, p. 738-741.

Lorrentz, P., Howells, W. G. J., & McDonald-Maier, K. D. (2010). A novel weightless artificial neural based multi-classifier for complex classifications. *Neural processing letters, 31*(1), 25-44.

Mafarja, M. M., Eleyan, D., Jaber, I., Hammouri, A., & Mirjalili, S. (2017). *Binary dragonfly algorithm for feature selection.* Paper presented at the 2017 International Conference on New Trends in Computing Sciences (ICTCS).

Malekzadeh, S., Gholizadeh, M. H., & Razavi, S. N. (2018). Persian Vowel recognition with MFCC and ANN on PCVC speech dataset. *arXiv preprint arXiv:1812.06953*.

Matsoukas, S., Gauvain, J.-L., Adda, G., Colthurst, T., Kao, C.-L., Kimball, O., . . . Makhoul, J. (2006). Advances in transcription of broadcast news and conversational telephone speech within the combined EARS BBN/LIMSI system. *Audio, Speech, and Language Processing, IEEE Transactions on, 14*(5), 1541-1556.

Matthews, B. A. (2012). *Probabilistic modeling of neural data for analysis and synthesis of speech.* Georgia Institute of Technology.

May, A., Garakani, A. B., Lu, Z., Guo, D., Liu, K., Bellet, A., . . . Kingsbury, B. (2017). Kernel approximation methods for speech recognition. *arXiv preprint arXiv:1701.03577*.

Mazenan, M. N., & Tan, T.-S. (2014). Malay Wordlist Modeling for Articulation Disorder Patient by Using Computerized Speech Diagnosis System. *Research Journal of Applied Sciences, Engineering and Technology, 7*(21), 4535-4540.

McLoughlin, I. (2009). *Applied speech and audio processing: with Matlab examples*: Cambridge University Press.

Mehla, R., & Aggarwal, R. (2014). Automatic Speech Recognition: A Survey. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 3*(1), pp: 45-53.

Mehraj, M., Butt, M. A., & Zaman, M. (2017). AUTOMATIC SPEECH RECOGNITION APPROACH FOR DIVERSE VOICE COMMANDS. *International Journal of Advanced Research in Computer Science, 8*(9).

Mengistu, K. T., & Rudzicz, F. (2011). *Comparing humans and automatic speech recognition systems in recognizing dysarthric speech.* Paper presented at the Canadian Conference on Artificial Intelligence.

Meutzner, H., Araki, S., Fujimoto, M., & Nakatani, T. (2016). *A generative-discriminative hybrid approach to multi-channel noise reduction for robust automatic speech recognition.* Paper presented at the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). *Applied multivariate research: Design and interpretation*: Sage publications.

Mishra, S., Bhowmick, A., & Shrotriya, M. C. (2016). Hindi vowel classification using QCN-MFCC features. *Perspectives in Science, 8,* 28-31.

Misra, D. D., Dutta, K., Bhattacharjee, U., Sarma, K. K., & Goswami, P. K. (2015). Assamese vowel speech recognition using GMM and

ANN approaches *Recent trends in intelligent and emerging systems* (pp. 163-170): Springer.

Moore, R., & Cunningham, S. (2005). *Plasticity in systems for automatic speech recognition: a review*. Paper presented at the ISCA Workshop on Plasticity in Speech Perception.

Morán-Fernández, L., Bolón-Canedo, V., & Alonso-Betanzos, A. (2017). Centralized vs. distributed feature selection methods based on data complexity measures. *Knowledge-Based Systems, 117*, 27-45.

Muche, M. (2020). *Development of an Amharic Speech to Ethiopian Sign Language Translation System.*

Müller, F. (2013). *Invariant Features and Enhanced Speaker Normalization for Automatic Speech Recognition*: Logos Verlag Berlin GmbH.

Mun, S., Park, S., Lee, Y., & Ko, H. (2016). Deep neural network bottleneck feature for acoustic scene classification. *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, 66*.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*: MIT press.

Musa, A. B. (2014). A comparison of ℓ1-regularizion, PCA, KPCA and ICA for dimensionality reduction in logistic regression. *International Journal of Machine Learning and Cybernetics, 5*(6), 861-873.

Nagraj, P., Jhawar, G., & Mahalakshmi, P. (2016). *Cepstral analysis of American vowel Phonemes (a study on feature extraction and comparison)*. Paper presented at the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET).

Najafabadi, M. M., Khoshgoftaar, T. M., & Seliya, N. (2016). Evaluating feature selection methods for network intrusion detection with kyoto data. *International Journal of Reliability, Quality and Safety Engineering, 23*(01), 1650001.

Nallasamy, U., Black, A. W., & Woszczyna, M. (2012). Acoustic and Lexical Modeling Techniques for Accented Speech Recognition.

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., & Turaga, D. S. (2017). *Learning Feature Engineering for Classification.* Paper presented at the IJCAI.

Nasereddin, H. H., & Omari, A. A. R. (2017). *Classification techniques for automatic speech recognition (ASR) algorithms used with real time speech translation.* Paper presented at the 2017 Computing Conference.

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE Access, 7*, 19143-19165.

Nguyen, K. (2016). *Adaptive Boosting for Automatic Speech Recognition*: Northeastern University.

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences, 1169*, 490-493.

Nicolao, M. (2019). *Context-aware speech synthesis: A human-inspired model for monitoring and adapting synthetic speech.* University of Sheffield.

Nivetha, S. (2020). *A Survey on Speech Feature Extraction and Classification Techniques.* Paper presented at the 2020 International Conference on Inventive Computation Technologies (ICICT).

Omar, N. M., & El-Hawary, M. E. (2017). *Feature fusion techniques based training MLP for speaker identification system.* Paper presented at the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).

OMer, D. (2020). *Amharic Named Entity Recognition System using Statistical Method.*

Palaz, D., Magimai-Doss, M., & Collobert, R. (2019). End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Communication, 108*, 15-32.

Pallant, J. (2013). *SPSS survival manual*: McGraw-Hill Education (UK).

Pandya, D. H., Upadhyay, S. H., & Harsha, S. P. (2014). Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. *Soft Computing, 18*(2), 255-266.

Pantano, E., & Vannucci, V. (2019). Who is innovating? An exploratory research of digital technologies diffusion in retail industry. *Journal of Retailing and Consumer Services, 49*, 297-304.

Parsons, S. (2010). Introduction to Machine Learning, Second Editon by Ethem Alpaydin, MIT Press, 584 pp., $55.00. ISBN 978-0-262-01243-0. *The Knowledge Engineering Review, 25*(03), 353-353.

Paulraj, M., Yaacob, S., & Yusof, S. M. (2008). *Vowel recognition based on frequency ranges determined by bandwidth approach.* Paper presented at the Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on.

Pedersen, C., & Diederich, J. (2007). *Accent classification using support vector machines.* Paper presented at the Computer and Information Science, 2007. ICIS 2007. 6th IEEE/ACIS International Conference on.

Penting, P. C. P., Rosdi, F., Mustafa, M. B., Salim, S. S., & Zin, N. A. M. (2019). Automatic Speech Intelligibility Detection for Speakers with Speech Impairments: The Identification of Significant Speech Features. *Sains Malaysiana, 48*(12), 2737-2747.

Phillips, J., Cripps, E., Lau, J. W., & Hodkiewicz, M. (2015). Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing, 60,* 316-325.

Phull, D. K., & Kumar, G. B. (2016). Vowel Analysis for Indian English. *Procedia Computer Science*, *93*, 533-538.

Plahl, C., Schlüter, R., & Ney, H. (2011). *Cross-lingual portability of Chinese and English neural network features for French and German LVCSR.* Paper presented at the Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on.

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713-731.

Prajapati, P., & Patel, M. (2017). A Survey on Isolated Word and Digit Recognition using Different Techniques. *International Journal of Computer Applications, 161*(3), 0975-8887.

Psutka, J. V., & Psutka, J. (2019). Sample size for maximum-likelihood estimates of Gaussian model depending on dimensionality of pattern space. *Pattern Recognition*, *91*, 25-33.

Qiao, F., Rosenfeld, R., & Sherwani, J. (2010). Layperson-Trained Speech Recognition for Resource Scarce Languages.

Rabiee, A., & Setayeshi, S. (2010). *Persian accents identification using an adaptive neural network.* Paper presented at the 2nd International Workshop on Education Technology and Computer Science.

Rabiner, L. R., & Schafer, R. W. (2009). Theory and application of digital speech processing. *Preliminary Edition*.

Radha, N., Shahina, A., & Khan, A. N. (2019). *Improving Recognition of Speech System Using Multimodal Approach.* Paper presented at the International Conference on Innovative Computing and Communications.

Rahman, F. D., Mohamed, N., Mustafa, M. B., & Salim, S. S. (2014). *Automatic speech recognition system for Malay speaking children.* Paper presented at the Student Project Conference (ICT-ISPC), 2014 Third ICT International.

Rahman, M. M., Dipta, D. R., & Hasan, M. M. (2018). *Dynamic time warping assisted svm classifier for bangla speech recognition.* Paper presented at the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2).

Rai, A., Khan, A., Bajaj, A., & Khurana, J. B. (2017). An efficient online examination system using speech recognition. *International Research Journal of Engineering and Technology, 4*(4), 2938-2941.

Rajaratnam, K., & Kalita, J. (2018). *Noise flooding for detecting audio adversarial examples against automatic speech recognition.* Paper presented at the 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).

Ramli, I., Jamil, N., Seman, N., & Ardi, N. (2015). An improved syllabification for a better Malay language text-to-speech synthesis (TTS). *Procedia Computer Science*, *76*, 417-424.

Ramteke, P. B., & Koolagudi, S. G. (2019). Phoneme boundary detection from speech: A rule based approach. *Speech Communication, 107,* 1-17.

Ratkevicius, K., Paskauskaite, G., & Bartisiute, G. (2018). Advanced recognition of Lithuanian digit names using hybrid approach. *Elektronika ir Elektrotechnika, 24*(2), 70-73.

Raveendran, V., Sanjeev, M. R., Paul, N., & Jijina, K. (2016). *Speech only interface approach for personal computing environment.* Paper presented at the 2016 IEEE International Conference on Engineering and Technology (ICETECH).

Razak, A. A., Abidin, M. I. Z., & Komiya, R. (2003). *Emotion pitch variation analysis in Malay and English voice samples.* Paper presented at the Communications, 2003. APCC 2003. The 9th Asia-Pacific Conference on.

Reid, C. V. (2016). *Study of Locus Equations as Features for Speech Classification and Recognition.* uga.

Reif, M., & Shafait, F. (2014). Efficient feature size reduction via predictive forward selection. *Pattern Recognition*, *47*(4), 1664-1673.

Retto, J. (2017). Sophia, first citizen robot of the world. *ResearchGate https://www. researchgate. net*.

Ribeiro, B. M., & Teixeira, C. A. D. (2017). *Finding the Critical Feature Dimension of Big Datasets.* Universidade de Coimbra.

Risanuri, H., Yulian, D., Bejo, A., & Sumaryono, S. (2017). *Comparison of vowel feature extraction on time and frequency domain.* Paper presented at the 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE).

Robert, C. (2014). Machine learning, a probabilistic perspective: Taylor & Francis.

Romera, M., Talatchian, P., Tsunegi, S., Araujo, F. A., Cros, V., Bortolotti, P., . . . Kubota, H. (2018). Vowel Recognition with Four Coupled Spin-Torque Nano-Oscillators. *Nature*, *563*(7730), 230-234.

Rottland, J., Neukirchen, C., & Willett, D. (1997). *Performance of hybrid MMI-connectionist/HMM systems on the WSJ speech database.* Paper presented at the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing.

Saa, J. F. D., & Gutierrez, M. S. (2010). *EEG signal classification using power spectral features and linear discriminant analysis: A brain computer interface application.* Paper presented at the Eighth Latin American and Caribbean Conference for Engineering and Technology.

Sahu, S., Mitra, V., Seneviratne, N., & Espy-Wilson, C. Y. (2019). *Multi-Modal Learning for Speech Emotion Recognition: An Analysis and Comparison of ASR Outputs with Ground Truth Transcription.* Paper presented at the Interspeech.

Sakai, J., & Doshita, S. (1962). The Phonetic Typewriter. *Information Processing*.

Saleema, A., & Thampi, S. M. (2018). Voice Biometrics: The Promising Future of Authentication in the Internet of Things *Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science* (pp. 360-389): IGI Global.

Salillari, D., & Prifti, L. (2016). A multinomial logistic regression model for text in Albanian language. *Journal of Advances in Mathematics, 12*(7), 6407-6411.

Salo, F., Nassif, A. B., & Essex, A. (2019). Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection. *Computer Networks, 148*, 164-175.

Sangwan, A., & Hansen, J. H. (2012). Automatic analysis of Mandarin accented English using phonological features. *Speech Communication, 54*(1), 40-54.

Sangwan, P., Deshwal, D., Kumar, D., & Bhardwaj, S. (2020). Isolated word language identification system with hybrid features from a deep belief network. *International Journal of Communication Systems*, e4418.

Sarkar, A. K., Do, C.-T., Le, V.-B., & Barras, C. (2014). Combination of cepstral and phonetically discriminative features for speaker verification. *IEEE Signal Processing Letters, 21*(9), 1040-1044.

Sarma, B. D., & Prasanna, S. M. (2018). Acoustic–phonetic analysis for speech recognition: A review. *IETE Technical Review, 35*(3), 305-327.

Scharenborg, O., Ciannella, F., Palaskar, S., Black, A., Metze, F., Ondel, L., & Hasegawa-Johnson, M. (2017). Building an asr system for a low-resource language through the adaptation of a high-resource language asr system: Preliminary results. *Proceedings of ICNLSSP, Casablanca, Morocco*.

Scharenborg, O., & Cooke, M. (2008). *Comparing human and machine recognition performance on a VCV corpus.* Paper presented at the Proc. Workshop on Speech Analysis and Processing for Knowledge Discovery.

Segura, J. C., Benítez, C., de la Torre, Á., Rubio, A. J., & Ramírez, J. (2004). Cepstral domain segmental nonlinear feature transformations for robust speech recognition. *IEEE Signal Processing Letters, 11*(5), 517-520.

Sen, S., Dutta, A., & Dey, N. (2019). Feature Extraction *Audio Processing and Speech Recognition* (pp. 45-66): Springer.

Shan, C., Weng, C., Wang, G., Su, D., Luo, M., Yu, D., & Xie, L. (2019). *Investigating end-to-end speech recognition for mandarin-english code-switching.* Paper presented at the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Sharan, R. V., & Moir, T. J. (2018). Pseudo-color cochleagram image feature and sequential feature selection for robust acoustic event recognition. *Applied Acoustics, 140*, 198-204.

Sharma, D., Wang, Y., Naylor, P. A., & Brookes, M. (2016). A data-driven non-intrusive measure of speech quality and intelligibility. *Speech Communication, 80*, 84-94.

Silva, J., Ribeiro, B., & Sung, A. H. (2017). *Finding the critical sampling of big datasets.* Paper presented at the Proceedings of the Computing Frontiers Conference.

Siraj, F., Azmi, M. Y. S., Paulraj, M., & Yaacob, S. (2009). *Malaysian Vowel Recognition Based on Spectral Envelope Using Bandwidth Approach.* Paper presented at the Modelling & Simulation, 2009. AMS'09. Third Asia International Conference on.

So, Y., & Cary, N. (1995). A tutorial on logistic regression. *SAS White Papers*.

Soltau, H., Saon, G., Kingsbury, B., Kuo, H.-K. J., Mangu, L., Povey, D., & Emami, A. (2009). Advances in Arabic speech transcription at IBM under the DARPA GALE program. *Audio, Speech, and Language Processing, IEEE Transactions on, 17*(5), 884-894.

Somervuo, P., Chen, B., & Zhu, Q. (2003). *Feature transformations and combinations for improving ASR performance.* Paper presented at the Eighth European Conference on Speech Communication and Technology.

Song, Y., Jiang, B., Bao, Y., Wei, S., & Dai, L.-R. (2013). I-vector representation based on bottleneck features for language identification. *Electronics Letters, 49*(24), 1569-1570.

Srinivas, N. S., Sugan, N., Kar, N., Kumar, L. S., Nath, M. K., & Kanhe, A. (2019). Recognition of Spoken Languages from Acoustic Speech Signals Using Fourier Parameters. *Circuits, Systems, and Signal Processing, 38*(11), 5018-5067.

Srinivasan, S., Fowler, S. P. J. E., Miller, J. B. L. T., Bruce, L., & Field, M. (2011). A Nonlinear Mixture Autoregressive Model for Speaker Verification.

Šuch, O., Barreda, S., Klimo, M., Tarábek, P., & Tinajová, A. (2018). *Comparing Classification Methods in Isolated Vowel Classification.* Paper presented at the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA).

Sudipa, I. N., Yadnya, I. B. P., Budiarsa, M., & Putra, I. N. D. (2018). Empowerment and preservation of local languages.

Sui, C., Bennamoun, M., & Togneri, R. (2017). Deep feature learning for dummies: A simple auto-encoder training method using Particle Swarm Optimisation. *Pattern Recognition Letters, 94*, 75-80.

Suktangman, N., Khanthavivone, K., & Songwatana, K. (2006). *Optimizing Vowel Recognition in Thai Spoken Language using Reduced LPC Spectrum and Reduced Feature Set of Critical Band Intensities.* Paper presented at the 2006 International Symposium on Communications and Information Technologies.

Suzuki, K., & Nakata, K. (1961). Recognition of Japanese Vowels—Preliminary to the Recognition of Speech. *J. Radio Res. Lab, 37*(8), 193-212.

Tabachnick, B., & Fidell, L. (2019). Using multivariate statistics 7th Edition: Pearson.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5): Pearson Boston, MA.

Tahir, M. A., Huang, H., Zeyer, A., Schlüter, R., & Ney, H. (2019). Training of reduced-rank linear transformations for multi-layer polynomial acoustic features for speech recognition. *Speech Communication, 110*, 56-63.

Tan, T. P., & Besacier, L. (2007). *Modeling context and language variation for non-native speech recognition.* Paper presented at the Interspeech.

Tan, T. P., Goh, S.-S., & Khaw, Y.-M. (2012). *A malay dialect translation and synthesis system: proposal and preliminary system.* Paper presented at the 2012 International Conference on Asian Language Processing.

Tan, T. P., & Ranaivo-Malançon, B. (2009). *Malay grapheme to phoneme tool for automatic speech recognition.* Paper presented at the Proc. Workshop of Malaysia and Indonesia Language Engineering (MALINDO) 2009.

Tan, T. P., Xiao, X., Tang, E. K., Chng, E. S., & Li, H. (2009). *MASS: A Malay language LVCSR corpus resource.* Paper presented at the 2009 Oriental COCOSDA International Conference on Speech Database and Assessments.

Teodorescu, H.-N. (2019). *How much are formants correlated and what that means?* Paper presented at the 2019 International Symposium on Signals, Circuits and Systems (ISSCS).

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern Recognition (4th edition)*: Academic Press.

Thiyagarajan, D., & Shanthi, N. (2019). A modified multi objective heuristic for effective feature selection in text classification. *Cluster Computing*, 1-11.

Tian, Y., Pei, K., Jana, S., & Ray, B. (2018). *Deeptest: Automated testing of deep-neural-network-driven autonomous cars.* Paper presented at the Proceedings of the 40th international conference on software engineering.

Ting, H. N., Chia, S. Y., Hamid, B. A., & Mukari, S. Z.-M. S. (2011). Acoustic characteristics of vowels by normal Malaysian Malay young adults. *Journal of Voice*, *25*(6), e305-e309.

Ting, H. N., & Lam, Y. (2009). *Speaker-Independent Malay Vowel Recognition of Children Using Neural Networks.* Paper presented at the World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany.

Ting, H. N., & Mark, K. (2008). *Speaker-dependent Malay Vowel Recognition for a child with articulation disorder using multi-layer perceptron.* Paper presented at the 4th Kuala Lumpur International Conference on Biomedical Engineering 2008.

Ting, H. N., Yong, B.-F., & Mirhassani, S. M. (2013). Self-adjustable neural network for speech recognition. *Engineering Applications of Artificial Intelligence, 26*(9), 2022-2027.

Ting, H. N., & Yunus, J. (2004). *Speaker-independent Malay vowel recognition of children using multi-layer perceptron.* Paper presented at the TENCON 2004. 2004 IEEE Region 10 Conference.

Ting, H. N., & Zourmand, A. (2011). *Gender Identification by Using Fundamental and Formant Frequency for Malay Children.* Paper presented at the 5th Kuala Lumpur International Conference on Biomedical Engineering 2011.

Tuncer, T., & Dogan, S. (2019). A novel octopus based Parkinson's disease and gender recognition method using vowels. *Applied Acoustics, 155*, 75-83.

Tyagi, P., & Shekhar, J. (2016). A Comparative Study of Multilayer Feed-forward Neural Network and Radial Basis Function Neural Network Models for Speech Recognition.

Unnibhavi, A. H., & Jangamshetti, D. (2017). *LPC based speech recognition for Kannada vowels.* Paper presented at the 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT).

Vadwala, A. Y., Suthar, K. A., Karmakar, Y. A., & Pandya, N. (2017). Survey paper on different speech recognition algorithm: Challenges and techniques. *Int. J. Comput. Appl., 175*(1), 31-36.

Vair, C., Colibro, D., Castaldo, F., Dalmasso, E., & Laface, P. (2006). *Channel factors compensation in model and feature domain for speaker recognition.* Paper presented at the 2006 IEEE Odyssey-The Speaker and Language Recognition Workshop.

Vaishnavi, V. K., & Kuechler, W. (2015). *Design science research methods and patterns: innovating information and communication technology*: Crc Press.

Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3): Cambridge university press.

Vieru, B., De Mareüil, P. B., & Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech Communication, 53*(3), 292-310.

Virkkunen, A. (2018). Automatic speech recognition for the hearing impaired in an augmented reality application.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2011). *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*: Springer Science & Business Media.

Wald, M., & Bain, K. (2008). Universal access to communication and learning: the role of automatic speech recognition. *Universal Access in the Information Society, 6*(4), 435-447.

Wang, M. (2011). *The classification of Mandarin finals using distinctive features.* Paper presented at the 2011 4th International Congress on Image and Signal Processing.

Wang, M., Zhang, J., Tang, H., Li, Z., Li, J., & Wang, Y. (2019). *Feature Analysis and Selection in Acoustic Events Detection.* Paper presented at the The International Conference on Cyber Security Intelligence and Analytics.

Wang, X., & Paliwal, K. K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition, 36*(10), 2429-2439.

Weisburd, D., & Britt, C. (2014). Multivariate regression with multiple category nominal or ordinal measures *Statistics in Criminal Justice* (pp. 601-636): Springer.

Williams, D., Escudero, P., & Gafos, A. (2018). Spectral change and duration as cues in Australian English listeners' front vowel categorization. *The Journal of the Acoustical Society of America, 144*(3), EL215-EL221.

Woo, S. C., Lim, C. P., & Osman, R. (2000). *Text-dependent speaker recognition using the fuzzy ARTMAP neural network.* Paper presented at the 2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No. 00CH37119).

Wrede, B. (2002). *Modelling the effects of speech rate variation for automatic speech recognition.* (Doktor-Ingenieurin Dissertation PhD), Universita¨t Bielefeld.

Wu, F., García-Perera, L. P., Povey, D., & Khudanpur, S. (2019). *Advances in Automatic Speech Recognition for Child Speech Using Factored Time Delay Neural Network.* Paper presented at the INTERSPEECH.

Wu, J., Yılmaz, E., Zhang, M., Li, H., & Tan, K. C. (2020). Deep spiking neural networks for large vocabulary automatic speech recognition. *Frontiers in neuroscience, 14*, 199.

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., . . . Zweig, G. (2016). Achieving Human Parity in Conversational Speech Recognition. *arXiv preprint arXiv:1610.05256*.

Yan, Q., Vaseghi, S., Rentzos, D., & Ho, C.-H. (2007). Analysis and synthesis of formant spaces of British, Australian, and American accents. *Audio, Speech, and Language Processing, IEEE Transactions on, 15*, 676-689.

Yeung, E. F. (2016). Applying Speech Recognition Methods to Analyze Singing.

Yong, B., & Ting, H. (2011). *Speaker-Independent Vowel Recognition for Malay Children using Time-Delay Neural Network.* Paper presented at the 5th Kuala Lumpur International Conference on Biomedical Engineering 2011.

Yu, D., & Deng, L. (2016). *Automatic Speech Recognition*: Springer.

Yu, D., & Seltzer, M. L. (2011). *Improved Bottleneck Features using Pretrained Deep Neural Networks.* Paper presented at the Twelfth annual conference of the international speech communication association.

Yuliani, A. R., Sustika, R., Yuwana, R. S., & Pardede, H. F. (2017). *Feature transformations for robust speech recognition in reverberant conditions.* Paper presented at the 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA).

Yusnita, M. (2014). *Investigation of Robust Speech Feature Extraction Techniques for Accents Classification of Malaysian English Speakers.* (Ph.D), Universiti Malaysia Perlis.

Yusnita, M., Paulraj, M. P., Yaacob, S., Shahriman, A. B., Yusuf, R., & Nordin, S. (2016). *On the Use of Spectral Feature Fusions for Enhanced Performance of Malaysian English Accents Classification.* Paper presented at the Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014).

Zahorian, S. A., Singh, T., & Hu, H. (2007). *Dimensionality Reduction of Speech Features using Nonlinear Principal Components Analysis.* Paper presented at the Eighth Annual Conference of the International Speech Communication Association.

Zamalloa, M., Rodríguez-Fuentes, L., Peñagarikano, M., Bordel, G., & Uribe, J. P. (2008). Feature Selection vs. Feature Transformation in Reducing Dimensionality for Speaker Recognition. *V Jornadas en Tecnología del Habla*, 45-48.

Zeki, A. M., & Azizah, N. (2001). *A Speech Synthesizer for Malay Language.* Paper presented at the National Conference on Research and Development in Computer Science (REDECS).

Zerari, N., Yousfi, B., & Abdelhamid, S. (2016). Automatic Speech Recognition: A Review. *Int. Acad. Res. J. Bus. Technol., 2*(2), 63-68.

Zhang, J., He, J., Wu, Z., Azmat, F., & Li, P. (2017). Prosodic Features-Based Speaker Verification using Speaker-Specific-Text for Short Utterances. *International Journal of Embedded Systems*, *9*(3), 250-261.

Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A. E.-D., Jin, W., & Schuller, B. (2018). Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments. *ACM Transactions on Intelligent Systems and Technology (TIST), 9*(5), 1-28.

Zhang, Z., Xu, S., Cao, S., & Zhang, S. (2018). *Deep Convolutional Neural Network with Mixup for Environmental Sound Classification.* Paper presented at the Chinese Conference on Pattern Recognition and Computer Vision (PRCV).

Zheng, D. C., Dyke, D., Berryman, F., & Morgan, C. (2012). A new approach to acoustic analysis of two British regional accents—

Birmingham and Liverpool accents. *International Journal of Speech Technology, 15*(2), 77-85.

Zhu, Q., Chen, B., Morgan, N., & Stolcke, A. (2004). *Tandem Connectionist Feature Extraction for Conversational Speech Recognition.* Paper presented at the International Workshop on Machine Learning for Multimodal Interaction.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands. *The Journal of the Acoustical Society of America, 33*, 248.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics : facts and models*: Springer-Verlag, Berlin.

Appendix A

Baseline Classification Rates

Classifiers	Malay vowels					Overall
	/a/	/e/	/i/	/o/	/u/	
ANN	98.29	96.93	99.56	91.94	91.37	95.62
MLR	98.00	97.82	99.71	94.43	92.42	96.48
LDA	97.64	94.06	97.99	90.14	88.17	93.60
KNN	88.39	88.40	90.27	86.48	72.69	85.25
Overall	95.58	94.30	96.88	90.75	86.16	92.74

Appendix B

Descriptive Statistics for Baseline MFCC FVs

Variable	Mean	Std. Error of Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis	Min.	Max.
MFCC1	-14.483	0.031	-14.524	1.383	1.912	.671	7.907	-18.230	1.336
MFCC2	4.914	0.025	5.211	1.111	1.235	-1.354	1.391	0.658	6.814
MFCC3	-0.867	0.031	-0.915	1.364	1.861	.139	-1.078	-3.650	2.173
MFCC4	1.929	0.028	1.658	1.236	1.527	.869	.122	-0.379	5.475
MFCC5	-0.458	0.017	-0.468	0.768	.590	.025	-.439	-2.298	2.259
MFCC6	-0.372	0.010	-0.348	0.426	.181	-.375	.288	-1.897	0.743
MFCC7	0.919	0.010	0.913	0.438	.192	.009	-.270	-0.426	2.370
MFCC8	-0.561	0.010	-0.535	0.459	.210	-.092	-.674	-1.897	0.607
MFCC9	0.304	0.009	0.288	0.384	.147	.210	.119	-0.830	1.500
MFCC10	-0.001	0.008	-0.027	0.359	.129	.137	-.093	-1.106	1.049
MFCC11	-0.228	0.007	-0.233	0.321	.103	-.013	-.376	-1.084	0.730
MFCC12	0.162	0.008	0.113	0.375	.141	.492	-.139	-0.692	1.371
MFCC13	-0.291	0.007	-0.287	0.329	.108	-.031	-.464	-1.243	0.676
MFCC14	-3.810	0.107	-3.379	4.739	22.460	-.468	1.103	-20.853	19.377
MFCC15	2.302	0.064	2.221	2.832	8.022	.180	2.916	-16.153	16.815
MFCC16	-0.760	0.052	-0.821	2.288	5.235	.076	1.049	-11.359	7.622
MFCC17	0.808	0.043	0.762	1.920	3.685	.268	1.831	-7.140	11.586
MFCC18	-0.623	0.035	-0.624	1.538	2.367	.371	1.803	-6.241	6.803

Variable	Mean	Std. Error of Mean	Median	Std. Deviation	Variance	Skewness	Kurtosis	Min.	Max.
MFCC19	-0.500	0.029	-0.472	1.285	1.652	.079	2.406	-6.056	8.676
MFCC20	0.588	0.030	0.608	1.311	1.720	-.312	2.355	-7.715	5.614
MFCC21	-0.524	0.030	-0.471	1.318	1.738	-.135	1.063	-6.048	5.115
MFCC22	0.253	0.028	0.247	1.259	1.584	.223	1.456	-3.982	7.002
MFCC23	0.104	0.023	0.161	1.027	1.054	-.102	.690	-3.855	4.486
MFCC24	0.234	0.021	0.221	0.942	.887	-.431	3.840	-5.648	4.183
MFCC25	0.035	0.023	0.007	1.007	1.015	-.071	1.824	-4.685	4.301
MFCC26	-0.099	0.023	-0.042	1.031	1.063	-.518	1.334	-4.423	3.309
MFCC27	0.341	0.057	0.391	2.503	6.263	-.195	3.153	-13.660	11.269
MFCC28	-1.557	0.052	-1.419	2.312	5.345	-.427	1.730	-11.466	7.593
MFCC29	1.211	0.038	1.113	1.662	2.762	.550	2.270	-4.854	10.922
MFCC30	-0.692	0.028	-0.610	1.247	1.556	-.863	3.167	-7.866	4.460
MFCC31	0.389	0.018	0.368	0.806	.650	.220	3.395	-3.671	5.411
MFCC32	0.316	0.018	0.241	0.804	.647	.713	4.382	-4.343	5.194
MFCC33	-0.457	0.021	-0.396	0.932	.868	-.506	4.030	-5.838	4.690
MFCC34	0.438	0.021	0.374	0.930	.864	.357	3.746	-5.177	5.199
MFCC35	-0.200	0.018	-0.211	0.806	.650	.256	5.010	-3.947	5.543
MFCC36	0.149	0.015	0.093	0.668	.447	.688	5.971	-4.180	5.117
MFCC37	-0.032	0.011	-0.032	0.505	.256	-.095	2.654	-2.617	2.428
MFCC38	-0.105	0.013	-0.086	0.557	.310	-.121	4.062	-4.525	3.109
MFCC39	0.058	0.012	0.072	0.541	.293	-.070	2.062	-2.551	2.456

Appendix C

Kolmogorov-Smirnov and Shapiro-Wilk Normality test for the Acoustic Features

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MFCC1	.041	1952	.000	.991	1952	.000
MFCC2	.164	1952	.000	.870	1952	.000
MFCC3	.070	1952	.000	.966	1952	.000
MFCC4	.116	1952	.000	.929	1952	.000
MFCC5	.034	1952	.000	.992	1952	.000
MFCC6	.038	1952	.000	.990	1952	.000
MFCC7	.018	1952	.115	.999	1952	.099
MFCC8	.032	1952	.000	.989	1952	.000
MFCC9	.031	1952	.000	.995	1952	.000
MFCC10	.035	1952	.000	.995	1952	.000
MFCC11	.025	1952	.006	.996	1952	.000
MFCC12	.055	1952	.000	.979	1952	.000
MFCC13	.029	1952	.001	.996	1952	.000
MFCC14	.068	1952	.000	.976	1952	.000
MFCC15	.047	1952	.000	.972	1952	.000
MFCC16	.027	1952	.002	.992	1952	.000
MFCC17	.052	1952	.000	.982	1952	.000
MFCC18	.054	1952	.000	.979	1952	.000
MFCC19	.047	1952	.000	.979	1952	.000
MFCC20	.045	1952	.000	.977	1952	.000
MFCC21	.034	1952	.000	.989	1952	.000
MFCC22	.044	1952	.000	.987	1952	.000

	Kolmogorov-Smirnov[a]	Shapiro-Wilk		Kolmogorov-Smirnov[a]	Shapiro-Wilk	
MFCC23	.031	1952	.000	.994	1952	.000
MFCC24	.056	1952	.000	.956	1952	.000
MFCC25	.040	1952	.000	.982	1952	.000
MFCC26	.052	1952	.000	.980	1952	.000
MFCC27	.053	1952	.000	.965	1952	.000
MFCC28	.051	1952	.000	.975	1952	.000
MFCC29	.061	1952	.000	.974	1952	.000
MFCC30	.076	1952	.000	.954	1952	.000
MFCC31	.053	1952	.000	.964	1952	.000
MFCC32	.089	1952	.000	.936	1952	.000
MFCC33	.073	1952	.000	.948	1952	.000
MFCC34	.067	1952	.000	.955	1952	.000
MFCC35	.074	1952	.000	.945	1952	.000
MFCC36	.100	1952	.000	.926	1952	.000
MFCC37	.063	1952	.000	.967	1952	.000
MFCC38	.060	1952	.000	.962	1952	.000
MFCC39	.043	1952	.000	.978	1952	.000

Appendix D

Test of Multicollinearity

Model	Pearson Correlation	Coefficients[a]	
		Collinearity Statistics	
	MFCC1	Tolerance	VIF
MFCC1	1.000	.106	9.441
MFCC2	-.024	.027	36.651
MFCC3	-.872	.058	17.217
MFCC4	-.166	.014	70.448
MFCC5	-.146	.044	22.911
MFCC6	.115	.063	15.934
MFCC7	-.117	.065	15.286
MFCC8	-.586	.051	19.635
MFCC9	-.209	.081	12.290
MFCC10	-.037	.092	10.856
MFCC11	.226	.125	7.992
MFCC12	.391	.128	7.803
MFCC13	-.117	.194	5.143
MFCC14	.203	.289	3.465
MFCC15	-.092	.179	5.587
MFCC16	-.039	.285	3.508
MFCC17	-.228	.260	3.849
MFCC18	-.299	.427	2.343
MFCC19	.162	.369	2.713
MFCC20	.087	.402	2.489
MFCC21	.146	.330	3.032
MFCC22	.010	.386	2.591
MFCC23	-.017	.455	2.197
MFCC24	.043	.455	2.196
MFCC25	.032	.460	2.174
MFCC26	-.045	.495	2.019
MFCC27	-.098	.315	3.176

MFCC28	.145	.111	9.040
MFCC29	-.002	.157	6.388
MFCC30	.152	.223	4.485
MFCC31	.185	.497	2.010
MFCC32	-.171	.422	2.368
MFCC33	.041	.302	3.310
MFCC34	-.132	.338	2.955
MFCC35	.008	.394	2.536
MFCC36	-.046	.508	1.968
MFCC37	-.115	.657	1.522
MFCC38	-.050	.673	1.486
MFCC39	.065	.630	1.587