Acoustic Comparison of Malaysian and Nigerian English Accents

A. F. Abdulwahab[1,2], S. A. Mohd Yusof[1] and H. Husni[1]

[1]*Institute of Advance and Smart Digital Opportunities*
[2]*School of Computing, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.*
abdulwahabfu@gmail.com

Abstract—**This study examines the differences in spectral and cepstral acoustics features between Malaysian and Nigerian English accents with the aim of determining the effect of accents on spectral and cepstral features of speech. Accent has received a great attention from ARS researchers due to the fact that it is a major source of ASR performance degradation. Most ASR applications were developed with native English speakers speech samples disregarding the fact that majority of its potential users speaks English as a second language with a marked accent. Malaysia and Nigeria were both colonized by Britain and speaks English as an official or second language despite being multi-ethnic nations. The results of the study revealed that formants values can be used to differentiate between ME and NE accents, most especially *F1* and *F2*. Cepstral (MFCC) performs better in accents recognition than formants features. While the combination of both formants and MFCC yields a better classification performance. However, the effect of the formants is non-uniform and depends on the vowels and accents under consideration. This is evident as each of the formants has different predictive values. Classification rate shows that Multi-Layer Perceptron (MLP) performs better than K-nearest neighbors (KNN).**

Index Terms—**Accent Recognition; Automatic Speech Recognition; Formants Analysis; KNN.**

I. INTRODUCTION

Communication plays a vital role in our daily life activities and interactions between fellow humans and machines as well. Out of the several means of communication such as writing, gesturing, posture, eye contact; speech is the most predominant means of human communication. Communication by speech involves verbal articulation, voicing, and fluency. Dissimilarities in the articulation of speech (sounds) have led to the emergence of several languages [1-3] with English being the most widely spoken language globally [4].

Several factors such as colonization, trade, tourism, and migration have significantly aided the spread of English to several regions of the world such as Africa, Asia and South America. These phenomena spread of English as expressed in [5] has given birth to different varieties of Englishes such as Nigerian English (NE), Malaysia English (ME), Singaporean English (SE) [6] resulting in English been spoken with diverse accents across the globe. Malaysia and Nigeria were both former colonies of British rule and also multi-ethnic nations with many ethnic groups that speak English with a unique accent that is dependent on their ethnic origin [7].

Although, the fact that ASR is fast becoming pervasive in our daily lives due to its deployment and applications such as in phone call voice dialing, interface to voice dictation and dialogue systems, navigation systems, biometric and authentication, Broadcast News transcription, speech control enable elevators and assistive aids for the elderly [8], which has brought conveniences to our daily living is not in doubt. Nonetheless, the major concern is about the high word error rate (WER) when ASR is exposed to accented speech. Although WER of ASR has been drastically reduced to less than 10% for few languages [9-12], the reverse is the case for most under-resourced languages. A high WER of 50% was recorded when ASR trained with American English (AE) was tested with NE speech data [13]. A similar test by [14] using six different regional accented English shows an average of 41.43% WER. This implies that accent variation constitutes a major source of performance degradation of ASR mainly attributable to the mismatch between the training set - native speech and testing set - non-native speech [15].

Several features such as temporal, prosodic, spectral, and cepstral exist within speech signals that provide information such as age, gender, emotion, accent, and health status of a speaker [16, 17]. Mel Frequency Cepstral Coefficient (MFCC) has found wider usage among ASR researchers due to its ability to mimic human's auditory system [16, 18-20]. Several research findings revealed that formants are a better indicator of accents, hence it's usage by accents researchers [16, 21]. While classifiers such as Support Vector Machine (SVM), Neural Network (NN), and K-Nearest Neighbors (KNN) are usually used for classification [16, 18]. It has been argued by numerous researchers that ability to correctly identify speaker's accent can greatly improve ASR performance [16-19, 22].This work shall focus on the English spoken accents of Malaysian and Nigerian English based on their formants and MFCC as features to be classified using Neural Network (NN) and K-Nearest Neighbors (KNN) classifiers to determine their differences and similarities if any, based ethnic origin.

II. RELATED WORK

Previous studies on accent have shown that the ability to correctly recognized accent has greatly enhanced the recognition performance of ASR when exposed to accented speech data. In a study of 14 regional accents of British using 19-MFCC, 12-Perceptual Linear Predictive (PLP) together with delta and double-delta features and Gaussian Mixture Model (GMM) and SVM as classifiers, [23] achieved a performance increase of 5.58%. A study by [14] using PLP features and GMM for 6 different regional accented English resulted in an average of 41.43% WER which was reduced to 27% on the incorporation of accent identification module. Several studies have explored numerous acoustic features of

speech such as energy, pitch, formants, MFCC, and LPC to establish the differences between regional or cross ethnics accent aimed at better understanding of the differences in the acoustic features to enhance ASR performance. Comparative analysis between the spectral acoustic features of British, Australian and American English accents was the focus of the study by [21] for the purpose of quantifying the differences between the three English accents. Results of the study revealed that formants are greatly affected by accent features. However, the effects are non-uniform across accents and phonemes. The results also revealed that formants are a better indicator of accents than MFCC features.

[13] established UISpeech corpus consisting of recordings from the three major ethnics of Nigeria – Hausa, Ibo and Yoruba for leveraging the ASR performance of a low resource language such as NE. Comparative analysis between NE and AE was carried out using acoustic parameters of the fundamental frequency ($F0$), formants ($F1$ and $F2$) and inter-Hidden Markov Model (HMM) distance extracted from UISpeech and TIMIT corpus in other to determine the differences between NE and AE. The result of the comparison shows that NE has a higher $F0$ value compare to AE. Likewise, AE has a higher value than the NE based on the formants space plot of $F1$-$F2$. Equally, the result of KL-divergence between AE and NE vowels reveals a distinct divergence between AE and NE pairs. Hence, it established the fact that there exist significant differences between AE and NE with a resultant effect on the poor performance of AE trained ASR when tested with NE data.

As argued by [17], the ability to correctly identified speaker's accent can significantly improve the performance of ASR in recognizing accented speech. In proofing their assertion, an experiment was conducted using speech samples from Marathi and Arabic speakers who read English digits 0 to 9. Extracted from the recorded speech database were acoustic features of energy, F0, $F1 – F5$. From the results of the analysis, it can be observed that Arabic-English accent has a higher energy value and also higher classification accuracy than Marathi English accent. Based on the classification accuracy, formant frequency, energy, and the pitch have the highest accuracy in that order for Marathi accents. While for Arabic accent, the order of accuracy is energy, formant frequency, and pitch. It can be deduced from their study, the pitch has the lowest correlation with accent, while formant frequency and energy produces dissimilar results for the two accents. The implication of this is that different acoustic features have different predictive values for dissimilar accents.

Similarly, [18] argued that accurate accent identification has the potent to enhance ASR performance. Classification experiment was conducted using KNN on the three accents of ME – Malay, Chinese, and Indian using acoustic features of LPC, log energy and formants. Based on the classification results of KNN classifier, formants $F1$ and $F2$ are significant for accent identification. This is followed by $F5$ while $F3$ and $F4$ have the least affect in accent identification. Similarly, recognition rates vary across the three accents for the different formant. In a study to identify Persian Accents by [16] using acoustic features of $F2, F3$ and 13-MFCC, classification results revealed that MLP performs better than SVM and KNN with an accuracy of 81% against 47%. Apparently, from the previous studies reviewed above, it is evident that accent constitutes an impediment to the performance of ASR. Hence, consequently serves as a barrier

to ASR wide reception and usage in real life situations. It therefore becomes pertinent that accent should be given adequate research attention with the view of enhancing ASR performance to accented speech which will inherently promote its wide acceptability and applicability globally.

III. EXPERIMENTAL SETUP

The experiment set up in this work consist of corpus formation, acoustic feature extractions and classification.

A. Speech Corpus

The speech corpus used in this study is made up of two separates corpus: NE and ME. The NE consist of 1500 utterances of five pure English vowels obtained from selected 30 Nigerian students from Universiti Utara Malaysia (UUM). The speakers are made up of 10 male from each of the major three ethnics of Nigeria - Hausa, Ibo and Yoruba with the average age 31. The ME corpus was obtained from the collection of [24]. The corpus consists of speech from Malay, Chinese and Indian male with a total of 694 utterances. Each of the speakers for both NE and ME corpus, read the 5 consonant-vowel (CV) pair of "KA", "KE", "KI", "KO", and "KU" representing five pure English vowels of /a/, /e/, /i/, /o/, and /u/ [18,19]. Each of the CV words was pronounced many times depending on the situation to improve the quality of the recordings. The details of the elicitation of the speech corpus used in this research is as given in Table 1 below.

As observed by [20], to mitigate the possible effect of smoking on voice quality, only non-smokers are selected for voice elicitation. The recordings were done is a relatively quiet room with a noise level of about 22dB which is considered normal [18]. The voices were recorded at 16 kHz for NE while ME were recorded at 8 kHz with a bit resolution of 16bps on a laptop using the software Audacity (Version 2.0.3) and Matlab respectively [21-23]. The recorded voices were saved as *.wav* format for further processing

Table 1
Speech Corpus Details

Accent/Settings	Gender	No of speakers.
Malaysian (ME)	Male	15
Nigerian (NE)	Male	30
Sampling frequency	16Khz, 8Khz	
Recording environment	Room	
Utterances recorded	"KA", "KE", KI", "KO", & "KU"	
Total no of utterances	2194	

B. Acoustic Features

From the corpus databases of both ME and NE, acoustic features of formants (F1-F5) and 39 MFCC were extracted from the pure vowels of English using Matlab codes. Due to the differences in the recording sampling frequency, the ME corpus is resampled from 8 to 16 kHz during feature extraction for uniformity of the data.

1) Formants (F1–F5)

Speech formants conveys vital information relating to accent characteristics and speaker identity [21]. Formants have received remarkable research efforts and become widely used features in ASR because it represents high concentrates of energy for voiced segment of speech. Formants are unique

in describing the phonetic nature of speech samples, most especially vowels [17], [25]. Basically, formants are made up of six frequencies depending on speaker characteristics and the phonemes, spanning a frequency range of 0-5kHz. The values of the formants are in increasing other, each higher than the preceding one. In this study, the first five formants denoted as F1, F2, F3, F4, and F5 respectively [26] were extracted from each of the pre-processed speech files using LPC roots method [27].

2) Mel-Frequency Cepstrum Coefficients (MFCCs)

MFCCs was developed by [25] and had since remained one of the most widely used features in ASR [26]. MFCCs are perceptually motivated speech representation which is based on Fourier discrete cosine transform of the log filter bank amplitudes. Modelled after human auditory system, MFCCs is built on Mel-frequency scale where each filter computes the average spectrum around each central frequency. MFCCs has been the most frequently used technique, especially in speech recognition and speaker verification applications. In addition to the regular 13-MFCC coefficients, we added to each of the 13 features cepstral features a delta, and a double delta or acceleration feature. Thereby making a total of 39-MFCC coefficients were extracted for classification purposes.

C. Classification

This study made use of two classifiers: Neural Network (MLP) and KNN to classify the features into five vowels classes based on accents. Both classifiers are trained and tested with randomized data of ratio 70% and 30% respectively. The classification rate (CR) is average of 10-fold training and testing for each of the classifiers.

1) Multi-Layer Perceptron (MLP)

A two-layer MLP was used to classify the features into five vowels classes. The number of neurons in the input layer is equal to the feature vector length which varies between 1 to 44. The network has 2 layers of 10 hidden neurons and 5 output neurons representing each of the vowels. The network is trained using Levenberg-Marquardt (LM) learning algorithm due to its fast convergence and accuracy [18], while mean-squared error (MSE) is used as an objective criterion for learning of the task.

2) The K-nearest neighbors (KNN)

KNN is a statistical prediction method in which an unknown pattern or query instance is predicted based on a simple popular vote of the categories or classes of the nearest neighbors in the training space. It works based on minimum distance from the query instance to the training samples to determine the K-nearest neighbors. Euclidean distance which is one of the popular methods being used as distance measure is used in this study with value of k = 2.

IV. FEATURES ANALYSIS AND ACCENT CLASSIFICATION

In this section, extracted formants mean values from both ME and NE corpus are compared. Subsequently, both formants and MFCC features are classified using NN (MLP) and KNN classifiers.

A. Formants Analysis – The mean values of formants (F1 – F5) for ME and NE are as shown in Table 2.

Table 2 shows the five formants values for five pure vowels of English as pronounced by Malaysians and Nigerians male. As evident form the table, formants values increases from *F1* up to *F5*. ME accents have a higher formants values for all the vowels except for formant *F5* where NE values is greater that of NE. This implies that ME and NE can be differentiated based on formants values.

Table 2
Mean values of formants for ME and NE vowels

Vowel/ Formant	/a/	/e/	/i/	/o/	/u/
ME					
F1	656.31	461.15	285.32	431.20	310.84
F2	1253.04	1437.62	1626.22	930.04	793.36
F3	2249.68	2398.20	2608.44	2408.95	2217.22
F4	3117.47	3309.35	3245.22	3317.42	3224.53
F5	3734.55	3845.06	3733.28	3889.58	3907.43
NE					
F1	531.01	324.00	242.31	350.62	265.15
F2	1084.33	1097.27	981.39	741.03	697.98
F3	1810.08	2279.31	2198.36	2034.20	1856.52
F4	2881.86	3049.54	3152.27	3116.12	2916.26
F5	3969.30	3938.01	3881.90	3997.86	4024.48

Table 3 presents the mean values of the five formants for NE and ME. The table shows the same trends as in the previous table 2 for vowels values. ME has higher formants values than NE except for *F5* in which NE has a higher formant value than NE. As evidenced from Figure 1, there exist a wider margin between the mean value of *F2* and *F3* of ME and NE. This suggests that ME and NE can be differentiated based on the mean of *F2* and *F3* values. These findings is similar to those of [16, 27]. Hence, using mean formants values especially *F2* and *F3,* ME and NE can be differentiated clearly.

Table 3
Mean values of formants for ME and NE Accents

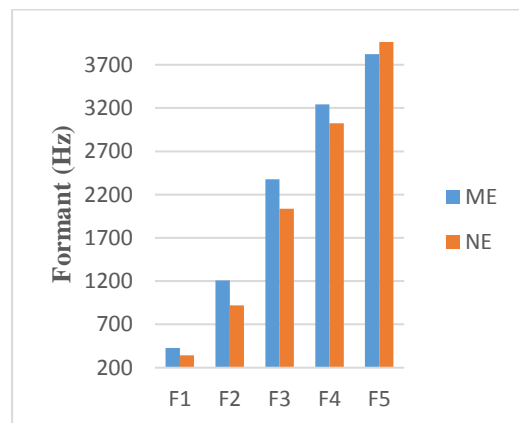| Accent	F1	F2	F3	F4	F5
ME	428.97	1208.06	2376.50	3242.80	3821.98
NE	342.62	920.40	2035.69	3023.21	3962.31

Figure 1: Average formants values for ME and NE

B. Classification

In this section, two classifiers of MLP and KNN are used to classify the formants and MFCC features. Several combinations of the features were combined and classified to determine their affinity or predictability on the accents.

1) Classification by single formants

The five formants values (F1-F5) extracted from both ME and NE were classified using MLP and KNN classifiers to determine the propensity for accent identification. The results of the classification are as shown in figure 2 below. From the figure 2, it shows that the classification performance is far below average using single formant value as the average classification rate (CR) is 40.2%. *F2* for ME accents gave the highest CR of 74.26% while *F5* has the least CR of 25.66% for MLP. *F2* also has the highest CR of 62.04%, while *F4* has the least CR of 20.41% for ME using KNN. For NE, *F1* has higher CR of 51.32% and *F5* has the least CR of 29.61% for MLP. Similarly, for KNN, *F1* has higher CR of 40.7% and *F5* has the least CR of 25.87% for NE. For ME, *F2* and *F1* perform better than the rest of the other formants. While for NE, *F1* and *F2* outperform other formants. On the overall, MLP classifiers perform better than KNN while ME has a better recognition rate than NE.

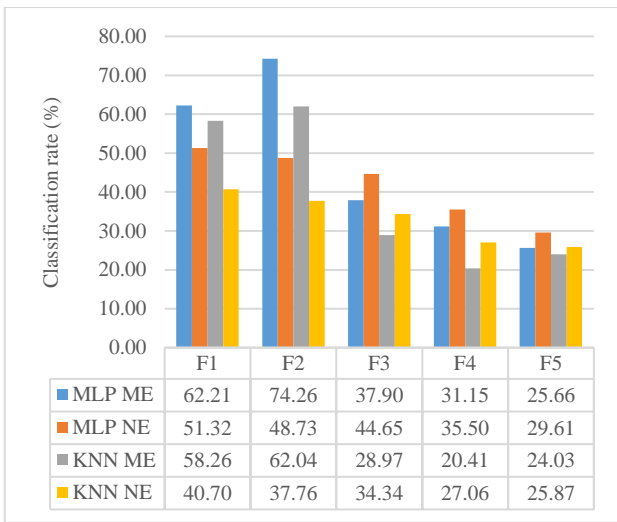|	F1	F2	F3	F4	F5
■ MLP ME	62.21	74.26	37.90	31.15	25.66
■ MLP NE	51.32	48.73	44.65	35.50	29.61
■ KNN ME	58.26	62.04	28.97	20.41	24.03
■ KNN NE	40.70	37.76	34.34	27.06	25.87

Figure 2: Classification rate of single formant value for ME and NE\

2) Classification by formants masking a formant value at a time

In order to determine which of the formants plays the significant role in accent identification, MLP and KNN classifiers are used to classify the five formants values (5F) and masking a formant value at a time for both ME and NE. The results of the classification is as shown in figure 3. The result shows improved CR of 84.95% as compared to 40.2% using single formant value as in figure 2. As expected, using all the five formants (5F) produces the best CR for both accents and classifiers. While masking a formant value at a time, resulting in different CR values based on the formants being masked and classifiers used. *F1* and *F2* by their lowest CRs for both MLP and KNN have the highest predictive value for ME accent with CR drop of 17.98% and 17.44% respectively. This is followed by *F4* and *F3,* while *F5* has the least effect with a drop of 3.39% in CR. For NE, *F1* and *F2* with a drop of 8.87% and 7.26% in CRs for both classifiers have the highest predictive value. This is followed by *F4* and *F3,* while *F5* has the least predictive value with a drop of 2.08% in CR. The implication of this is that each of the five formants has various levels of predictive values to the accents. Hence, identification of their predictive value can help in accent identification and subsequent improvement of ASR performance when faced with accented speech.
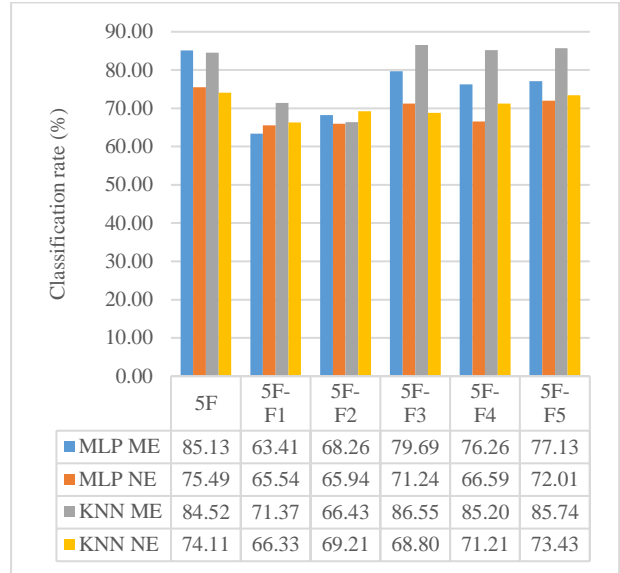

	5F	5F-F1	5F-F2	5F-F3	5F-F4	5F-F5
■ MLP ME	85.13	63.41	68.26	79.69	76.26	77.13
■ MLP NE	75.49	65.54	65.94	71.24	66.59	72.01
■ KNN ME	84.52	71.37	66.43	86.55	85.20	85.74
■ KNN NE	74.11	66.33	69.21	68.80	71.21	73.43

Figure 3: Classification rate for formants masking a format at a time for ME & NE

3) Classification by 39-MFCC features

In this experiment, MLP and KNN classifiers are used to classify 39-MFCC features extracted from both ME and NE data. The CRs is as shown in figure 4 below. CRs shows that MFCC features have a better CR than using formant values. The best CR is attained by KNN for vowel /i/ of ME with 99.2%. Surprisingly the same KNN achieved the least CR for vowel /u/ of NE with 68.99%. Based on the average CRs, MLP performs better than KNN. While for accents, ME has a better CR of 92.1% as against NE with CR of 84.81%. This result is however contrary to the findings by [21] that formants are stronger indicators of accents than cepstrum features.
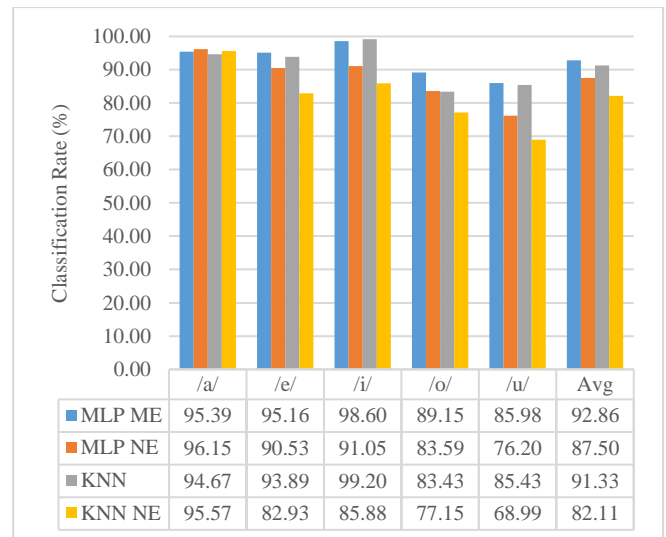

	/a/	/e/	/i/	/o/	/u/	Avg
■ MLP ME	95.39	95.16	98.60	89.15	85.98	92.86
■ MLP NE	96.15	90.53	91.05	83.59	76.20	87.50
■ KNN	94.67	93.89	99.20	83.43	85.43	91.33
■ KNN NE	95.57	82.93	85.88	77.15	68.99	82.11

Figure 4: Classification Results for 39-MFCC for ME & NE

4) Classification by 39-MFCC and formants features

Lastly, 39-MFCC together with different combinations of formants values were classified using MLP and KNN classifiers. The results of the classification are as shown in figure 5 below. When compared with the CRs of 39-MFCC in figure 3, the CRs however shows a divergent result for the two classifiers. The combination of 39-MFCC and the 5F

shows marginal improvement of 1.66% and 1.2% for ME and NE respectively using MLP. While the combination of 39-MFCC with 5F features using KNN result in a drop of CR of -7.42% and -9.44% for ME and NE respectively. This shows that CRs is dependent on the features and classifier used. Based on masking each of the formants, the predictive values of the formants are in the following order for ME and NE accents is F1, F3, F2, F4, and F5. This is contrary to the results obtained in figure 2 where F1 and F2 are the most significant formants. For the classifiers, MLP performs better than KNN. While for accents, ME has a better recognition rate than NE.

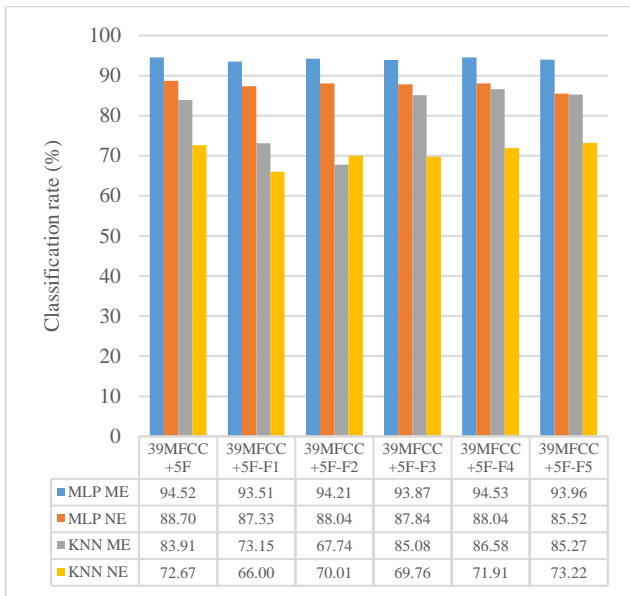|	39MFCC +5F	39MFCC +5F-F1	39MFCC +5F-F2	39MFCC +5F-F3	39MFCC +5F-F4	39MFCC +5F-F5
MLP ME	94.52	93.51	94.21	93.87	94.53	93.96
MLP NE	88.70	87.33	88.04	87.84	88.04	85.52
KNN ME	83.91	73.15	67.74	85.08	86.58	85.27
KNN NE	72.67	66.00	70.01	69.76	71.91	73.22

Figure 5: Classification Results for 39-MFCC plus formants for ME & NE

V. CONCLUSION

In this paper, comparative analysis of formants and 39-MFCC features of ME and NE is carried out. The mean value of formants is higher for ME than NE except for *F5*. Accents classification by formants features yields below average CR of 40.2%, while using 5F resulted in 100% increase in CR over single formant. Masking each of the formants at a time reveals that *F1* and *F2* have the most predictive value and *F5* the least for both ME and NE. 39-MFCC features gave better CRs than the formants. However, a marginal CRs improvement was attained using a combination of 39-MFCC and formants features. From the mean average values of the five formants and classification outputs of both MLP and KNN, its reveals that there exists a significant difference between acoustic values of ME and NE accents. Based on these features ME and NE can be differentiated, especially using *F1* and *F2* values. We also show that unique features have different predictability values for the accents. Based on the classifiers, MLP performs better than KNN for both accents.

REFERENCES

[1] S. Furui, "Fifty years of progress in speech and speaker recognition," *Journal of the Acoustical Society of America,* vol. 116, no. 4, pp. 2497-2498, 2004.

[2] M. Hariharan, L. S. Chee, O. C. Ai, and S. Yaacob, "Classification of speech dysfluencies using LPC based parameterization techniques," *Journal of Medical Systems,* vol. 36, no. 3, pp. 1821-1830, 2012.

[3] C.-L. Huang and C.-H. Wu, "Spoken document retrieval using multilevel knowledge and semantic verification," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 15, no. 8, pp. 2551-2560, 2007.

[4] D. Crystal, *English as a global language*: Cambridge University Press, 2012.

[5] J. Foley, *English in new cultural contexts: Reflections from Singapore*. Oxford University Press, 1998.

[6] S. Sharbawi, *An Acoustic Investigation of the Segmental Features of Educated Brunei English Speech.* Nanyang Technological University, English Language and Literature, 2009.

[7] U. Gut, "Nigerian English: Phonology," in *A handbook of varieties of English,* vol. 1, 2004, pp. 992-1002.

[8] U. Nallasamy, F. Metze, and T. Schultz, "Active learning for accent adaptation in automatic speech recognition," in *2012 IEEE Spoken Language Technology Workshop (SLT),* 2012, pp. 360-365.

[9] J.-L. Gauvain, G. Adda, M. Adda-Decker, A. Allauzen, V. Gendner, L. Lamel, and H. Schwenk, "Where are we in transcribing French broadcast news?," in *INTERSPEECH*, 2005, pp. 1665-1668.

[10] S. Matsoukas, J.-L. Gauvain, G. Adda, T. Colthurst, C.-L. Kao, O. Kimball, L. Lamel, F. Lefevre, J. Z. Ma, and J. Makhoul, "Advances in transcription of broadcast news and conversational telephone speech within the combined EARS BBN/LIMSI system," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 14, pp. 1541-1556, 2006.

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

[12] R. Hsiao, M. C. Fuhs, Y.-C. Tam, Q. Jin, and T. Schultz, "The CMU-interACT 2008 Mandarin transcription system," in *INTERSPEECH*, 2008, pp. 1445-1448.

[13] S. Amuda, H. Boril, A. Sangwan, and J. H. Hansen, "Limited resource speech recognition for Nigerian English," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2010, pp. 5090-5093.

[14] D. Vergyri, L. Lamel, and J.-L. Gauvain, "Automatic speech recognition of multiple accented English data," in *INTERSPEECH*, 2010, pp. 1652-1655.

[15] R. Soorajkumar, G. Girish, P. B. Ramteke, S. S. Joshi, and S. G. Koolagudi, "Text-Independent Automatic Accent Identification System for Kannada Language," in *Proceedings of the International Conference on Data Engineering and Communication Technology*, 2017, pp. 411-418.

[16] A. Rabiee and S. Setayeshi, "Persian accents identification using an adaptive neural network," in Proceedings of the *2nd International Workshop on Education Technology and Computer Science*, 2010, pp. 7-10.

[17] S. Gaikwad, B. Gawali, and K. Kale, "Accent Recognition for Indian English using Acoustic Feature Approach," *International Journal of Computer Applications,* vol. 63, 2013.

[18] M. Yusnita, M. P. Paulraj, S. Yaacob, S. A. Bakar, and A. Saidatul, "Malaysian English accents identification using LPC and formant analysis," in Proceedings of the *2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE),* 2011, pp. 472-476.

[19] C. Pedersen and J. Diederich, "Accent classification using support vector machines," *in Proceedings of the 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)*, 2007, pp. 444-449.

[20] X. Huang, "Australian Accent-Based Speaker Classification," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 416-419.

[21] S. Ghorshi, S. Vaseghi, and Q. Yan, "Cross-entropic comparison of formants of British, Australian and American English accents," *Speech Communication,* vol. 50, pp. 564-579, 2008.

[22] L. M. Arslan and J. H. L. Hansen, "Language accent classification in American English," *Speech Communication,* vol. 18, pp. 353-367, 1996.

[23] A. Hanani, M. J. Russell, and M. J. Carey, "Human and computer recognition of regional accents and ethnic groups from British English speech," *Computer Speech & Language,* vol. 27, pp. 59-74, 2013.

[24] S. Azmi, F. Siraj, S. Yaacob, M. Paulraj, and A. Nazri, "Improved Malay Vowel Feature Extraction Method Based on First and Second Formants," *in Proceedings of the 2nd International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM),* 2010, pp. 339-344.

[25] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously

spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing, ,* vol. 28, pp. 357-366, 1980.

[26] M. Yusnita, "Investigation of Robust Speech Feature Extraction Techniques for Accents Classification of Malaysian English Speakers,"

Ph.D Thesis, School of Mechatronic Engineering, Universiti Malaysia Perlis, 2014.

[27] L. Arslan and J. Hansen, "A study of temporal features and frequency characteristics in American English foreign accent,"*Journal of the Acoustical Society of America,* vol. 102, pp. 28-40, 1997.