ACCENT IDENTIFICATION OF MALAYSIAN AND NIGERIAN ENGLISH BASED ON ACOUSTIC FEATURES

[1]Abdulwahab F. Atanda, [2]Shahrul Azmi Mohd. Yosuf & [3]H. Husni

[1] *Institute of Advance & Smart Digital Opportunities (IASDO)*,

[1,2,3]*School of Computing, Universiti Utara Malaysia (UUM)*

Sintok, UUM 06010 Kedah, Malaysia

abdulwahabfu@gmail.com

[2]*shahrulazmi@uum.edu.my*

[3]*husniza@uum.edu.my*

ABSTRACT

Purpose - This paper studies acoustics features of energy, pitch and formants of Malaysian and Nigerian English vowels with the aim of effective accents identification using multi liner regression (MLR) and linear discriminant analysis (LDA) classifiers for performance improvement of ASR when exposed to accented speech. Accent being a foremost source of ASR performance degradation has received a great attention from ASR researchers. Majority of ASR applications were developed with native English speakers speech samples without considering fact that most of its potential users speaks English as a second language with a marked accent, hence its poor performance when exposed to accented speech. Previous studies on accent has 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, (Hanani, Russell, & Carey, 2013) achieved a performance increase of 5.58%. A study by (Vergyri, Lamel, & Gauvain, 2010) using six different regional accented English shows an average of 41.43% WER. Which was reduced to 27% on the incorporation of accent identification module. Several studies have explored several 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. Apparently from the previous studies reviewed above, it is evident that accent constitute a hurdle to the performance of ASR. Hence, consequently serves as a barrier to ASR wide reception and usage in real life situations. Consequently, it becomes pertinent that accent should be given adequate research attention with the view of enhancing ASR performance to accented speech which will inherently promotes its wide acceptability and applicability globally.

Analysis shows that ME occupies more formant space than NE. Eo, F2, F3 and F4 are significant in differentiating the two accents. ME has a higher CR than NE, while MLR performs better than LDA. Fusion of the features performs better than any of the individual acoustic features.

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

1. Dataset- The speech dataset used in this study is made up of two separate dataset: ME and NE. The ME dataset was obtained from the collection of (MY, Siraj, Yaacob, Paulraj, & Nazri, 2010). The dataset consist of speech from Malay, Chinese, and Indian male. 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 of 10 male from each of the major three ethnics of Hausa, Ibo, and Yoruba. Each of the speakers for both NE and ME dataset, read the 5 consonant-vowel (CV) pair of “KA”, “KE”, “KI”, “KO”, and “KU” representing five vowels of /a/, /e/, /i/, /o/, and /u/ (MY et al., 2010).

2. Acoustic Features- From the speech dataset, acoustic features energy, pitch and formants (F1-F5) and were extracted from the pure vowels of English using Matlab codes.

3. Classification-Accent identification is done using two classifiers: MLR and LDA by classifying the features into five vowels classes based on ME and NE accents. Both classifiers are trained and tested with randomized data of ratio 70% and 30% respectively. While training and testing is done with 10-folds cross validation.

Findings - Analysis of the result shows that, NE accent has a higher Eo and F0 value than ME. However, while the difference in their Eo values is significant enough to differentiate between the accent, the difference of their F0 value is insignificant. For the case of formants, F1 – F4 values of ME is higher than NE. NE has a higher F5 value than ME. Based on the differences in formants values, F2, F3, and F4 have significant differential values that can be explored to differentiate between the accents. F5 and F1 have the least differential values between the accents. Figure 1 below gave a graphical representation of the relationship between the acoustic features and the accents.

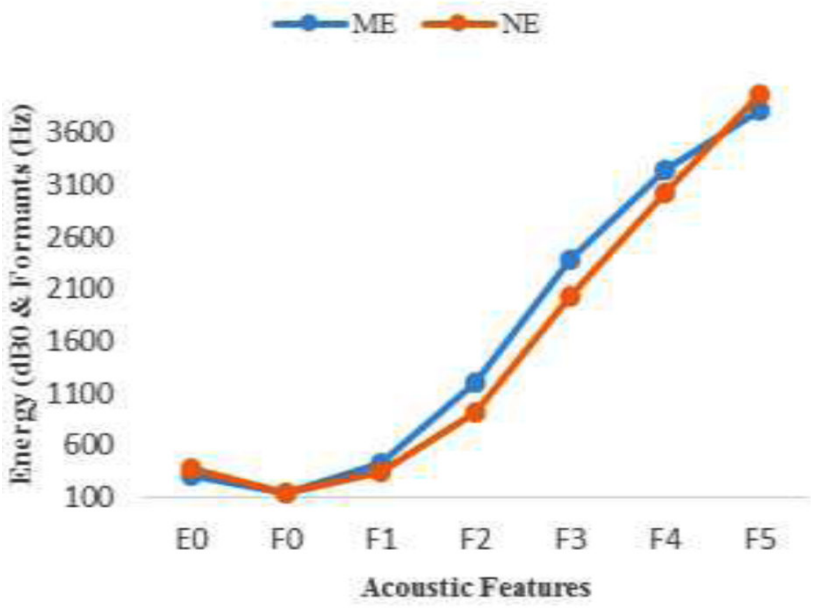

Figure 1: Average formants values for ME & NE

Keywords: Automatic speech recognition; Accent recognition; Multi Liner Regression (MLR) ; Linear Discriminant Analysis (LDA); Formants analysis.

CONCLUSIONS

In this study, we carried out vowel analysis of ME and NE accents. Acoustic features of Eo, F0, F1, F2, F3, F4, and F5 were extracted from data sets of Malaysian and Nigerian males. Analysis of EFFs shows that NE has higher values of Eo, F0 and F5. ME has higher values in F1, F2, F3, and F4. The values of EFFs are also distinct for each of the vowels across the accents. We conducted accent identification of the two accents using MLR and LDA classifiers based on the different combinations of the above acoustic features. Based on EFFs CR, vowel /a/ has the highest CR and vowel /i/ has the least CR value. ME has higher CR than NE. In the future, we intend to explore other acoustic features such as MFCC and LPC together with their derivatives and explore different combinations and masking of features.

REFERENCES

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.

MY, S. A., Siraj, F., Yaacob, S., Paulraj, M., & Nazri, A. (2010). *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.

Vergyri, D., Lamel, L., & Gauvain, J.-L. (2010). *Automatic speech recognition of multiple accented English data.* Paper presented at the INTERSPEECH.