Acoustic Analysis of Nigerian English Vowels Based on Accents

A F Atanda[1,2], S A Mohd Yusof[2], and H Husni[2]
[1]*Institute of Advance & Smart Digital Opportunities (IASDO).*
[2]*School of Computing Universiti Utara Malaysia (UUM), Sintok, Kedah, Malaysia.*
abdulwahabfu@gmail.com

***Abstract*—Accent has been widely acclaimed to be a major source of automatic speech recognition (ASR) performance degradation. Most ASR applications were developed with native English speaker speech samples not minding the fact that the majority of its potential users speaks English as a second language with a marked accent. Nigeria like most nations colonized by Britain, speaks English as official language despite being a multi-ethnic nation. This work explores the acoustic features of energy, fundamental frequency and the first three formats of the three major ethnic groups of Nigerian based on features extracted from five pure vowels of English obtained from subjects who are Nigerians. This research aimed at determining the differences or otherwise between the pronunciations of the three major ethnic nationalities in Nigeria to aid the development of ASR that is robust to NE accent. The results show that there exist significant differences between the mean values of the pure English vowels based on the pronunciation of the three major ethnics: Hausa, Ibo, and Yoruba. The differences can be explored to enhance the performance of ASR in recognition of NE.**

***Index Terms*—Accent Recognition; Acoustic Analysis; Automatic Speech Recognition; Formant Analysis; Nigeria English.**

I. INTRODUCTION

Communication is an essential and the most effective means of human interactions. The most predominant means of human interaction is speech. Speech is a verbal means of communication that entails an articulation, voice, and fluency. Differences in the articulation of speech (sounds) led to the emergence of several languages [1] - [3] with English being the most widely spoken language globally [4]. However, the majority of English speakers speaks English as a second (ESL) or foreign language with accents different from that of the native speakers.

The effect of colonization, trade, and migration has aided the spread of English to other parts of the world such as Africa, Asia and South America. The phenomena spread of English as expressed in [5] has given birth to different varieties of Englishness such as Nigerian English (NE), Malaysia English (ME), Singaporean English (SE) resulting in English being spoken with several accents across the globe [6]. Nigeria being a multi-ethnic nation with over 400 different ethnic groups speaks English with a unique accent that is dependent on ethnic origin [7]. This work shall focus on the English spoken accents of the three major ethnics of Hausa (H), Ibo (I) and Yoruba (Y) [8] based on the acoustic features of energy, fundamental frequency and the first-three formants values to determine their differences and similarities if any, based on ethnic origin.

Robustness of the human auditory system (HSR) have enabled it to recognize speech with high accuracy irrespective of speaker's characteristics such as age, gender, and to adapt quickly to variations such as accent and/or environmental situations such as noise [9], [10]. The quest to replicate this impressive high recognition and fast adaptation ability of HSR in machines motivated the development of ASR systems. ASR is basically a user interface for converting spoken words into text and actions. Advancement in technology has made speech recognition technology an indispensable tool for socioeconomic development and assistive technology.

Although ASR technology has witnessed appreciable advancement since its debut in the 1950's, ASR performance is however far below that of HSR with an error rate as high as 45% when ASR is exposed to non-native speakers (foreign accent) [10], [11], [13]. This degradable performance of ASR is attributed to non-cognizance of variabilities such as accent in a real-world situation during ASR design [10]. As [13] remarks that ASR systems are highly susceptible to speaker variability and that aside gender, the next source of variability is speech is accent and consequently suggested that ASR should be designed considering variation in accents rather than base on native speakers alone. Accent been a major source of variation that degrades and consequently constitute a big challenge to ASR performance [14], [15] as compared to the human ability to adapt and recognize speech spoken with different accents and in a different context. This thus calls for serious attention to accent as viable means of achieving ASR robustness.

II. PREVIOUS STUDIES

Previous studies on accent revealed that ability to accurately recognized accent has substantially improved the recognition performance of ASR when subjected to accented speech data. In a study of 14 regional accents of British, [14] attained a performance increase from 89.6% to 95.18%. A study by [15] using six different regional accented English shows an average of 41.43% WER. This was reduced to 27% on the incorporation of accent identification module.

UISpeech corpus made up of recordings from the three major ethnics of Nigeria – Hausa, Ibo and Yoruba were established by [16]. Acoustic parameters of fundamental frequency (F0), formants (F1 and F2) and inter-*HM*M distance were extracted from UISpeech corpus in other to determine the differences between NE and American English (AE). The analysis of the values reveals that NE has a higher F0 value as compared to the AE. The plot of F1-F2 reveals that AE has a higher value than the NE counterparts.

Likewise, the result of KL-divergence between AE and NE vowels shows a clear divergence between AE and NE pairs. Hence, it concludes that there exist substantial differences between AE and NE with a consequent effect on the poor performance of AE trained ASR when exposed to NE. However, this study does not perform a comparative analysis based on different ethnic groups constituting the NE to determine their similarities or otherwise.

It was asserted by [17] that the ability to accurately identified speaker's accent can greatly improve the performance of ASR in recognizing accented speech. To proof their assertion, an experiment was carried out using speech samples from Marathi and Arabic speakers who read English digits 0 to 9. Acoustic features of energy, F0, F1, F2, F3, F4, and F5 were extracted from the recorded speech for analysis. The results of the analysis show 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 inferred from the study that pitch has the lowest affinity with an accent, and that formant frequency and energy gives different results for the two accents. This implies that different acoustic features have different predictive values for different accents.

Also arguing the case of accent identification to enhance ASR performance, [18] experimented on the three accents of Malaysian English (ME) – Malay, Chinese, and Indian using acoustic features of LPC, log energy and formants. Of the 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. Also, recognition rates vary across the three accents for the different formant. As evident from the foregoing, accent constitute a barrier to the performance of ASR. Hence constitute a hindrance to ASR wide acceptance and application in real world situations. It is, therefore, pertinent that accent should be given adequate research attention to enhancing ASR performance and applicability globally.

III. EXPERIMENTAL SETUP

In conducting the experiment involved in this work, two processes of corpus formation and acoustic feature extractions were carried out as follows:

A. *Speech Corpus*

The speech corpus for this study is composed of 3,000 utterances of five pure English vowels obtained from selected 60 Nigerians who are students of Universiti Utara Malaysia (UUM). The speakers are made of 10 males and 10 females from each of the major three ethnics of Hausa, Ibo, and Yoruba. The average age of the female speakers is 29 and 31 for the male speakers. The overall average age for the speakers is 30. Each of the speakers read the 5 consonant-vowel (CV) pair of "KA", "KE", "KI", "KO", and "KU" representing five English vowels of /a/, /e/, /i/, /o/, and /u/ [18], [19]. Each of the CV words was pronounced 10 times to improve the quality of the recordings. Prior to recording, each of the subject were informed of the motive for the recording and also mock recordings were done to familiarize the subjects to the actual recording. Table 1 below gives the details of the elicitation of the speech corpus used in this research. In Table 2 shows the vowels used in this research together with their IPA notation is displayed

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 22 dB which is considered normal [18]. The voices were recorded at 16 kHz with a bit resolution of 16 bps on a laptop using the software Audacity (Version 2.0.3) [21]-[23]. The recorded voices were saved as .wav format for further processing.

Table 1
Speech Corpus Details

Accent/Settings	Gender	No of speakers.	No of utterances per speakers
Hausa (H)	Female	10	50
	Male	10	50
Ibo (I)	Female	10	50
	Male	10	50
Yoruba (Y)	Female	10	50
	Male	10	50
Sampling frequency	16khz		
Recording environment	Room		
Recorded utterances	KA", "KE", "KI", "KO" and "KU"		
Total no of utterances	3000		

Table 2
Phonetic symbols representation

Phone	a	e	i	o	u
IPA	Λ	e	i	o	u

B. *Acoustic features*

From the total of 3,000 utterances collected, we extracted acoustic features of E0, F0, F1, F2, and F3 from the pure vowels of English as described below using Matlab codes.

Energy (E0) - Energy is an important feature of speech with the potential to distinguish different accents from each other and also differentiates between languages. Energy being a good correlate of phoneme identity is a valuable cue to phoneme detection [24]. Given a speech sample of frame x, the energy in the frame is the sum over of the power of all the samples in the frame. Thus, energy is a speech frame x, can be estimated using energy equation given below.

$$E0 = \sum_{t=t_1}^{t_2} x^2[t] \tag{1}$$

where:
x = is the sample frame of the speech signal
t_1 = starting time of the signal frame
t_2 = ending time of the signal frame.

Fundamental Frequency (F0) – Fundamental frequency represents a unique feature of speech that is widely used in ASR, most especially in gender adaptive ASR [17]. Human speech perception is highly dependent on cues from F0. Based on pitch contours, differences in accent is visible, hence several researches make use of pitch in combination with other features for speech recognition [19].

Formants (F1, F2, and F3) – Formants has become widely used features in ASR due to the fact that it represents high concentrates of energy for voiced segment of speech. Formants are very vital in defining the phonetic nature of speech samples [17], [25]. Vowels formants are known to be highly indicative of accents. Structurally, formants are made up of six frequencies each higher than the preceding one. First, second, and third formants denoted as F1, F2, and F3 respectively [26] were extracted from each of the pre-processed speech files using LPC roots [27].

IV. Statistical Analysis

We conducted a two-way ANOVA analysis between the accents (CC) and extracted acoustic features (Predictors) to determine the predictability of the accents by the acoustic features. As shown in Table 3, $p < 0.05$ (.000) shows that the regression model statistically significantly predict the outcome variable of accents. Hence, we use all the five acoustic features (E0, F0, F1, F2, and F3) for further analysis in this study.

Having statistically determined the significance of the acoustic features based on the result of ANOVA test, we estimate statistical means (average) of acoustic features: E0, F0, F1, F2, and F3 using SPSS package. Table 4 below gives the statistical mean of acoustic features for the vowels extracted from the corpus of NE based on gender (F for female and M for male) and average (Avg) - combination of both genders. The mean values displayed is for each of the five vowels used in this research work.

Based on the mean statistical values of acoustic features as shown in Table 4, the following deduction can be made:

Energy (E0) – For vowels, /a/, /e/ and /o/, the mean E0 values of the female is higher than that of the male. While male have higher E0 value in vowels /i/ and /u/. Overall, female have higher E0 value than the male. Vowel /e/ has the highest E0 value on the average. While vowel /u/ has the least average value of E0. The overall average mean E0 value for NE vowels ranges from 272 – 469 dB.

Table 3
Two-way ANOVA results showing the significant value of accents (cc) and acoustic features (predictors)

ANOVA[a]						
	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	560.681	5	112.136	56.006	.000[b]
	Residual	1189.319	594	2.002		
	Total	1750.000	599			

a. Dependent Variable: CC
b. Predictors: (Constant), *F3*, *F0*, energy, *F1*, *F2*

Table 4
Mean values of acoustic features of NE vowels

Vowel	E0			F0			F1			F2			F3		
	F	M	Avg	F	M	Avg	F	M	Avg	F	M	Avg	F	M	Avg
/a/	351	338	344	201	139	170	496	531	513	994	1084	1039	1690	1810	1750
/e/	519	419	469	198	134	166	413	324	368	1099	1097	1098	2359	2279	2319
/i/	379	403	391	175	137	156	290	242	266	862	981	922	2139	2198	2169
/o/	466	386	426	188	140	164	414	351	382	834	741	787	1985	2034	2010
/u/	257	288	272	207	144	175	340	265	303	769	698	733	1812	1857	1834

Fundamental frequency (F0) – As expected, the mean F0 of females is higher than that of males for all the vowel values being considered. This shows that the acoustic value of F0 is unique for gender identification. This implies that performance of ASR can be improved by incorporating gender identification module using F0 value as a unique value. Average F0 values for females ranges between 175-207 Hz. For males, the average FO values ranges between 134-144 Hz. The average mean value of F0 for NE is between 156-175 Hz.

The mean value of F1 for the vowel /a/ in males is higher than that of females. The female has a higher F1 in the remaining vowels /i/, /e/, /o/ and /u/. Like F0, F1 mean value is also equally unique for gender identification. This implies that performance of ASR can be improved by incorporating gender identification module using F1 mean value as a unique identifier. The average mean value of F1 for NE is between 266-513 Hz.

F2 mean value for females is higher in vowels /e/, /o/ and /u/, while males have higher value than female in vowels /a/ and /i/. Average F2 values for males is between 698-1097 Hz. While for females, the average F2 values ranges between 769-1099 Hz. Mean value for F2 of NE ranges between 733-1098 Hz.

For F3, male has a higher mean value in vowels /a/, /i/, /o/ and /u/, while females have a higher value in vowel /e/. This implies that the value of F3 can be used to distinctly identify the vowels based on gender. On the overall, the mean value of F3 for NE is between 1750 – 2319 Hz. F3 might not be significant for gender discrimination.

As evident from Table 4, the statistical mean of acoustic features of vowels /a/, /e/, /i/, /o/ and /u/ are unique and different for both females and males. Exploring the differences in these values to design gender adaptive ASR can effectively improve ASR performance considerably as argued by [19]. Figure 1 (a – e) is a graphical representation of Table 4 that shows the mean values of acoustic features of E0, F0, F1, F2 and F3 for vowels /a/, /e/, /i/, /o/, and /u/ based on gender. Figures 1(a-c) clearly shows that there exist significant differences between the acoustic values of E0, F0 and F1 of NE for both males and females. On the contrary, figures 1(d-e) indicated that there exist insignificant differences between acoustic values of F2 and F3 of NE for both males and females. We therefore concluded that gender adaptive ASR can de designed based of the acoustic features of E0, F0, and F1 for NE.

Figure 1(f) shows the average acoustic values of NE for the pure English vowels. It gives the average values of vowels based on the acoustic features. E0 has the least value and F3 has the highest value. Vowel /e/ has a consistent higher value for all the acoustic features considered in this work. This is followed by vowel /i/ while vowel /u/ has the least value.

To determine the differences or similarities between vowel pronunciations by the three ethnics of NE consider in this research in relation to their gender, the mean values of each ethnic/gender and overall mean for each gender were obtained as displayed in Table 4. Figure 2 (a – e) gives a graphical representation of the means of the vowels under study.

Based on the mean statistical values of acoustic features in Table 4 below, vowel /e/ has the highest E0 value of 781.05 Hz. Vowel /u/ has the least E0 value of 221.47 Hz. Based on ethnic mean value, Ibo female (IF) has the highest E0 value in vowel /e/ while Yoruba females (YF) have the least value Vowel /u/. In terms of F0, vowel /u/ has the highest F0 value of 248.85 Hz while vowel /o/ has the least F0 value of 128.27 Hz.

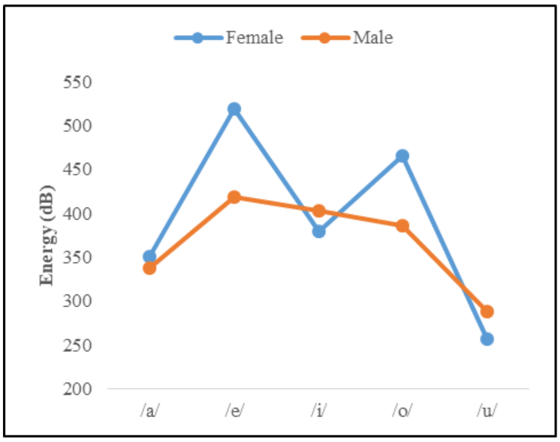


Figure 1(a): Mean E0 value of NE English vowels

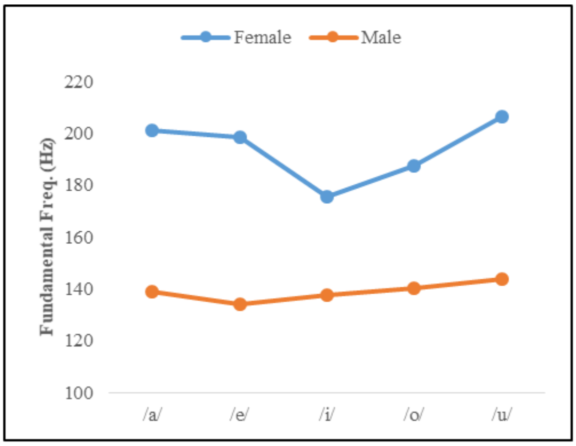


Figure 1(b): Mean Pitch value of NE for pure English vowels

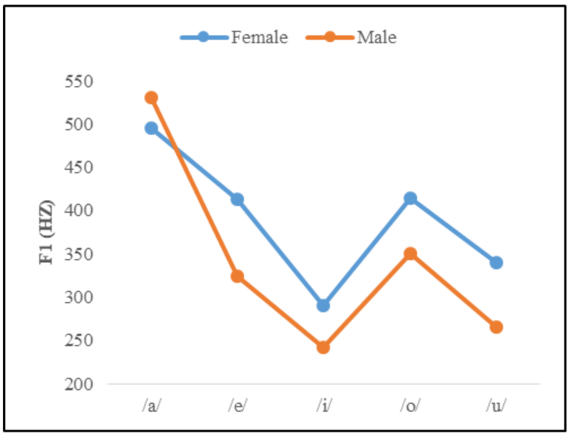


Figure 1(c): Mean F1 value of NE English vowels

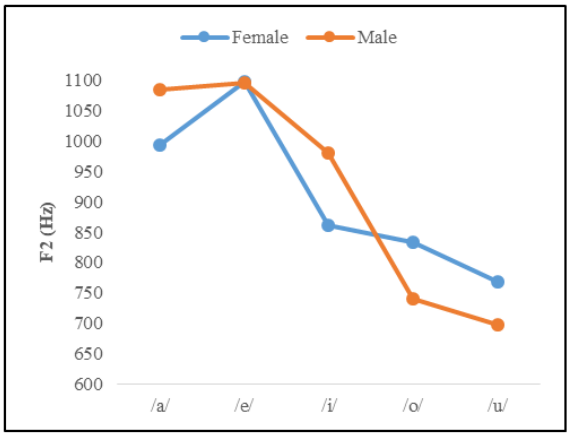


Figure 1(d): Mean F2 value of NE for pure English vowels

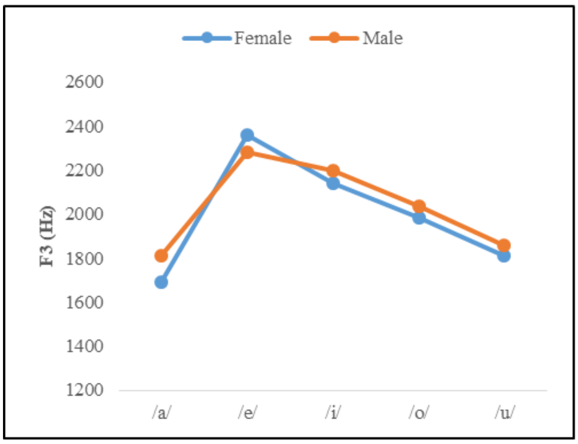


Figure 1(e): Mean F1 value of NE English vowels

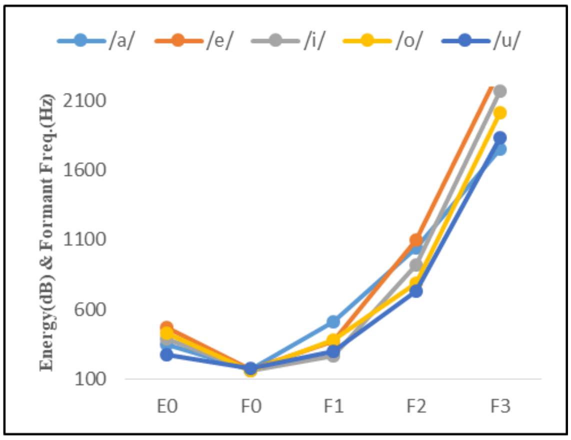


Figure 1(f): Mean Energy value of NE for pure English vowels

Table 4
Mean values of acoustic features of NE vowels

vowel a	HF	IF	YF	HM	IM	YM	H	I	Y
E0	283.70	524.39	245.00	336.48	391.33	285.25	310.09	457.86	265.13
F0	208.10	200.16	194.72	148.65	135.64	132.84	178.38	167.90	163.78
F1	459.88	511.18	515.97	445.77	586.14	561.12	452.83	548.66	538.55
F2	919.75	1085.26	975.88	986.11	1161.02	1105.88	952.93	1123.14	1040.88
F3	1833.56	1594.22	1642.22	1798.29	1823.14	1808.81	1815.93	1708.68	1725.52

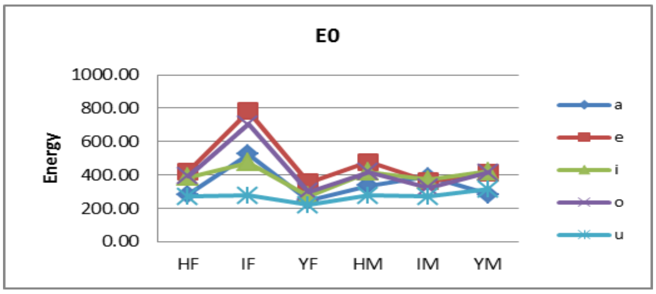
vowel e	HF	IF	YF	HM	IM	YM	H	I	Y
E0	421.78	781.05	354.90	476.81	362.97	416.67	449.29	572.01	385.79
F0	174.33	211.28	209.87	136.77	130.80	134.39	155.55	171.04	172.13
F1	379.28	461.23	397.51	313.05	335.54	323.42	346.16	398.39	360.47
F2	1284.40	969.26	1043.53	1115.31	1043.87	1132.62	1199.85	1006.56	1088.08
F3	2447.28	2311.17	2320.00	2273.60	2262.87	2301.45	2360.44	2287.02	2310.72

vowel i	HF	IF	YF	HM	IM	YM	H	I	Y
E0	386.16	475.03	275.40	417.99	371.54	420.82	402.07	423.28	348.11
F0	175.11	196.43	154.49	143.12	136.48	132.69	159.12	166.45	143.59
F1	261.85	297.39	310.68	232.04	243.77	251.12	246.94	270.58	280.90
F2	747.40	881.99	957.53	1060.76	1036.87	846.53	904.08	959.43	902.03
F3	1954.94	2395.35	2066.29	2175.72	2162.72	2256.64	2065.33	2279.03	2161.47

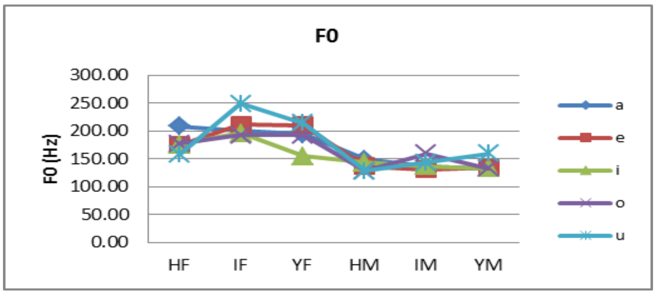
vowel o	HF	IF	YF	HM	IM	YM	H	I	Y
E0	395.00	704.25	299.47	418.28	323.48	415.02	406.64	513.86	357.24
F0	177.60	192.49	192.58	128.46	159.46	133.35	153.03	175.98	162.97
F1	374.04	490.06	377.74	328.90	366.15	356.82	351.47	428.10	367.28
F2	790.25	901.11	809.27	724.10	784.34	714.64	757.18	842.72	761.95
F3	1963.79	2061.95	1930.06	2067.43	2016.21	2018.96	2015.61	2039.08	1974.51

vowel u	HF	IF	YF	HM	IM	YM	H	I	Y
E0	270.37	278.63	221.47	278.03	269.95	315.54	274.20	274.29	268.50
F0	157.20	248.85	213.65	128.27	143.94	159.62	142.73	196.40	186.63
F1	307.85	392.07	321.49	248.51	274.12	272.82	278.18	333.10	297.16
F2	742.92	809.96	754.04	728.64	732.46	632.84	735.78	771.21	693.44
F3	1934.71	1634.45	1866.31	1883.86	1817.61	1868.09	1909.29	1726.03	1867.20

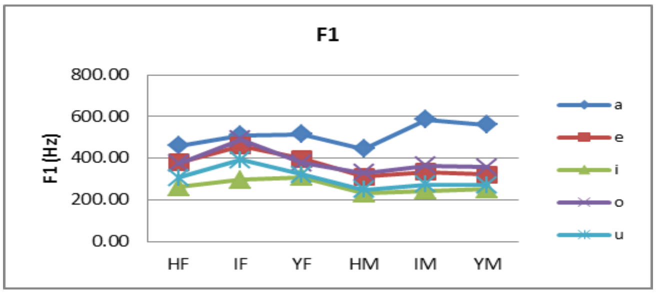
Based on ethnic mean value, Ibo females (IF) have the highest F0 value while Hausa males (HM) have the least value. As expected, the mean F0 of females is higher than that of males for all the vowel values being considered. This shows that the acoustic value of F0 is unique for gender identification. This implies that performance of ASR can be improved by incorporating gender identification module using F0 value as a unique value. For formants, F1 value of vowel /a/ is the highest with 586.14 Hz. Vowel /i/ has the least value of 232.04 Hz. Based on gender and ethnicity, Ibo male (IM) has the highest F1 value while HM has the least. The mean value of F2 for vowel /e/ is the highest with 1284.4 Hz. Vowel /u/ has the least value of 632.84 Hz. H female (HF) has the highest value of F2 of all the ethnic while YM has the least value of F2. For F3, vowel /e/ has the highest F3 value of 2447.28 Hz, while vowel /a/ has the least F3 value of 1594.22 Hz. HF average value for F3 is the highest among the ethnics of NE considered, likewise IF has the least F3 value. As evident from Table 4, the statistical mean of acoustic features for the vowels /a/, /e/, /i/, /o/ and /u/ are unique and different for both female and male and also across the ethnics. Exploring the differences in these values to design both gender and accent adaptive ASR can effectively improve ASR performance considerably as argued by [19]. In terms of differences of E0, vowels /e/ and /o/ showed a difference of more than 20% between males and females whereas vowels /a/ and /i/ only gave a difference of less than 7%. Other than F0, Formants F1 and F2 values can be also be used to classify gender due to its significant differences for vowels /i/, /o/ and /u/. Figure 2 (a – e) show the graphical plot of acoustic features of the three ethnics in terms of gender and an overall average of NE means values against the five pure English vowels.



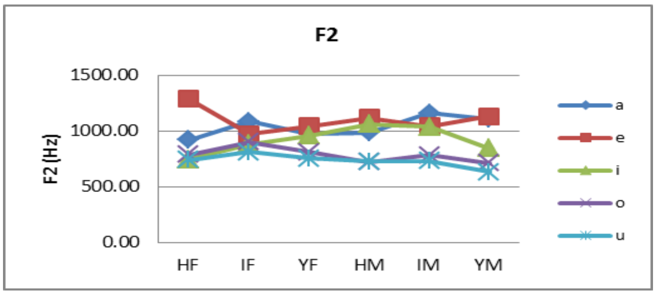
(a)



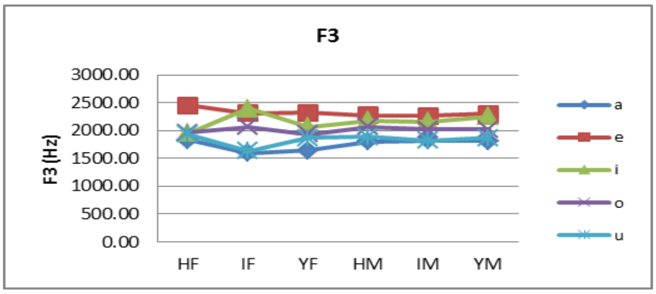
(b)



(c)



(d)

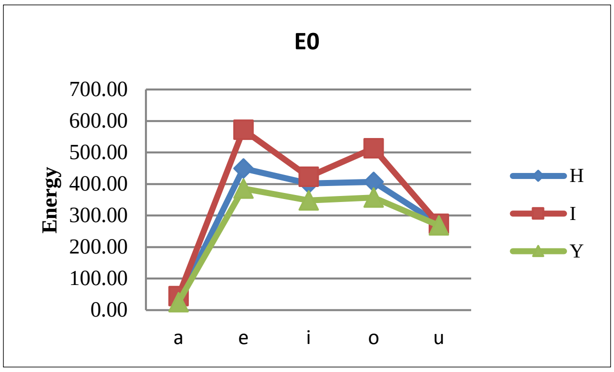


(e)

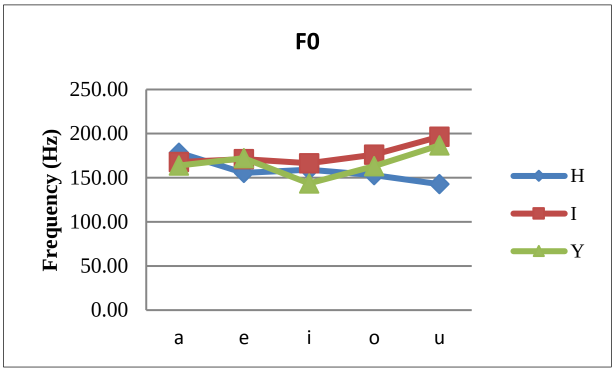
Figure 2: Graphical plot of acoustic features of the three ethnics in terms of gender and an overall average of NE means values against the five pure English vowels

From the plot of Figure 3 (a – e) which shows the mean values of acoustic features of NE vowels and the three ethnics group of Hausa, Ibo and Yoruba, the followings inference can be made.

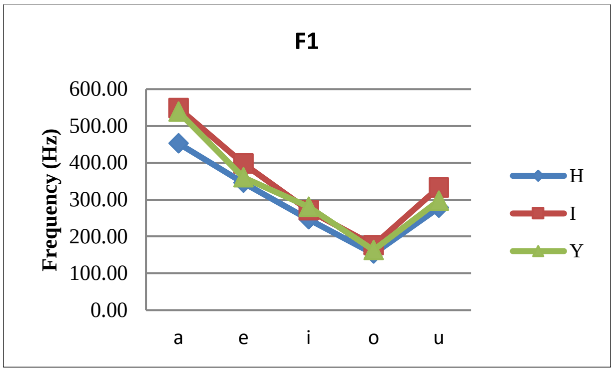
Considering the E0 values, Ibo accent has the highest E0 value followed by Hausa, while Yoruba has the least value. The E0 value for vowel /e/ is the highest while vowel /a/ has the least value. As evident is Fig. 3 (a), the E0 values for the ethnics are distinct for each of the vowels, hence mean of E0 can be used to recognize each of the accents. As for F0, Ibo accent has the highest value, followed by Yoruba medium, while Hausa has the least F0 value. Though the mean values of F0 of the vowels are different, however the different is blurring for a clear distinction except for the value of vowel /u/ that is significantly different. This suggests that the three accents can be differentiated based on the F0 value of vowel /u/. Similarly, the same observations can be made for formants (F1 – F3) values where there are no significant differences between the mean values of the vowels for the three accents. Nonetheless, unique differences in some of the vowels can be explored as a means of differentiation. For F1, the mean values vowel /a/ and /u/ can be used to identify each of the accents. Likewise, for F2, vowel /e/ is unique for identifying the accents. Equally, vowel /i/ and /u/ values for F3 can uniquely be explored for accents identification of NE. It can be inferred that based on the acoustic features examined, the three accents has distinctive means values indicating that there are significant differences based on pronunciation. In terms of classifying ethnics of Hausa (***H***), Ibo (***I***) and Yoruba (***Y***), ***F1*** and ***F2*** values for all vowels can be used as features to differentiate the ethnics.



(a)

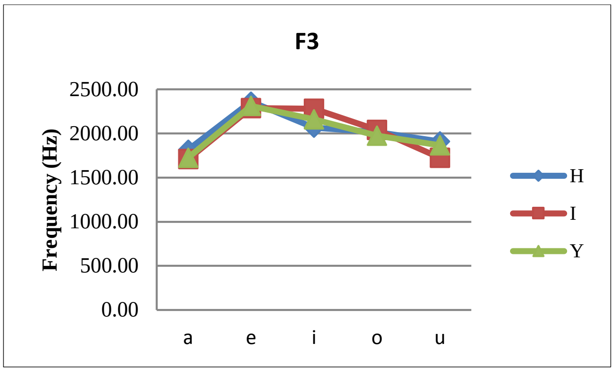


(b)



(c)

(d)



(e)

Figure 3: Mean values of acoustic features of NE vowels and the three ethnics group of Hausa, Ibo and Yoruba

V. CONCLUSION

In this paper, statistical analysis of mean of acoustic features of energy, fundamental frequency, and the first three formants was determined to establish the differences or otherwise between the three NE accents of Hausa, Ibo and Yoruba. The mean values of all the acoustic feature considered revealed that there exist significant differences in the acoustic values of the three accents studied across all the five pure vowels of English. The observed differences in the mean values for the vowels based on the three accents indicated that the acoustic differences can be explored in accent identification of NE. The results also reveal that F0 values for the gender differ significantly as in the previous research. This implies that gender identification can be effectively done by the values of F0, F1 and even F2 features. Further analysis shall be carried out in the future based on other acoustic feature to establish the differences in the accents of three ethnics of NE based on the pronunciation. One interesting finding from this study is E0 and formant (F1 and F2) features can significantly help to classify gender and ethnics depending on the vowels used.

REFERENCES

[1] Furui S. Fifty years of progress in speech and speaker recognition. *The Journal of the Acoustical Society of America*. 2004;116:2497.

[2] Hariharan M, Chee LS, Ai OC, Yaacob S. Classification of speech dysfluencies using LPC based parameterization techniques. *Journal of medical systems*. 2012;36:1821-30.

[3] Huang C-L, Wu C-H. Spoken document retrieval using multilevel knowledge and semantic verification. *Audio, Speech, and Language Processing, IEEE Transactions on*. 2007;15:2551-60.

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

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

[6] Sharbawi SH. *An acoustic investigation of the segmental features of educated Brunei English speech*, 2010.

[7] Gut U. *Nigerian English: Phonology. A handbook of varieties of English*. 2004;1:992-1002.

[8] Olaniyi OK. The taxonomy of Nigerian varieties of spoken English. *International Journal of English and Literature*. 2014;5(9):232-40.

[9] Scharenborg O, Cooke M, editors. Comparing human and machine recognition performance on a VCV corpus. *Proc Workshop on Speech Analysis and Processing for Knowledge Discovery*; 2008.

[10] You H, Adviser-Alwan *A. Robust automatic speech recognition algorithms for dealing with noise and accent*: University of California at Los Angeles; 2009.

[11] Lippmann RP., Speech recognition by machines and humans. *Speech Communication*. 1997;22:1-15.

[12] Stern RM, Morgan N., *Hearing Is Believing: Biologically--Inspired Feature Extraction For Robust Automatic Speech Recognition*. 2012.

[13] Faria A., Accent classification for speech recognition. *Machine Learning for Multimodal Interaction*: Springer; 2006. p. 285-93.

[14] Hanani A, Russell MJ, Carey MJ., Human and computer recognition of regional accents and ethnic groups from British English speech. *Computer Speech & Language*. 2013;27:59-74.

[15] Vergyri D, Lamel L, Gauvain J-L, editors., Automatic speech recognition of multiple accented English data. *INTERSPEECH*; 2010.

[16] Amuda SAY, Boril H, Sangwan A, Ibiyemi TS, Hansen JH. Engineering Analysis and Recognition of Nigerian English: An Insight into Low Resource Languages. *Transactions on Machine Learning and Artificial Intelligence*. 2014;2(4):115-28.

[17] Gaikwad S, Gawali B, Kale K., Accent Recognition for Indian English using Acoustic Feature Approach. *International Journal of Computer Applications*. 2013;63(7).

[18] Yusnita M, Paulraj MP, Yaacob S, Bakar SA, Saidatul A, editors., Malaysian English accents identification using LPC and formant analysis. *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*; 2011: IEEE.

[19] Azmi MS., *Development of Malay Word Pronunciation Application using Vowel Recognition*. Malay. 2016;9(1).

[20] Yusnita M., *Investigation of Robust Speech Feature Extraction Techniques for Accents Classification of Malaysian English Speakers [Ph.D]*: Universiti Malaysia Perlis; 2014.

[21] Evans J, Chu M-n, Aston JA, Su C-y., Linguistic and human effects on F0 in a tonal dialect of Qiang. *Phonetica*. 2010;67:82-99.

[22] Ghai W, Singh N., Literature Review on Automatic Speech Recognition. *International Journal of Computer Applications*. 2012;41(8):42-50.

[23] Hao Y-C., Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics*. 2012;40:269-79.

[24] Jurafsky D, Martin JH., *Speech and language processing*: Pearson; 2014.

[25] Shrawankar U, Thakare VM., *Techniques for feature extraction in speech recognition system: a comparative study*. arXiv preprint arXiv:13051145. 2013.

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

[27] Rabiner LR, Schafer RW., *Theory and application of digital speech processing*. Preliminary Edition. 2009.