JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 10(4), DECEMBER, 2022
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com

Development of an English Accent Database for Speech Analysis and Recognition

Djagbe Sirbele, Habib Rabi'u
Department of Electrical Engineering
Bayero University, Kano.

ABSTRACT
The multi ethnic nature of Nigeria and its strong attachments to values and traditions makes it unique among other African nations. The three main ethnic groups namely, Hausa, Igbo and Yoruba can in most cases be distinguished in terms of the attires they use and the style of their dressing or more generally by their physique. Nonetheless, socialization and intermarriages are making these factors fast disappearing. Yet, another physiological characteristic of humans is found in their sound production which contains interesting features that can be used to classify individual ethnicity. However, lack of locally developed audio database has hindered research progress in this area, unlike in other cultures were marked language database is in abundance. This research is focus on developing multi ethnic-based audio database that could be used for research in the areas of speech analysis, voice recognition and automatic answering machine system among others. The developed database contained 12,960 short documented words recorded from 72 participants drawn across the three major ethic groups namely Hausa, Ibo and Yoruba. The recordings were done in ultra-modern oral documentation studio using Audacity recording software (v 2.3.2).

ARTICLE INFO
Article History
Received: June, 2022
Received in revised form: July, 2022
Accepted: September, 2022
Published online: December, 2022

KEYWORDS
Audio database, English accent, ethnic group, isolated words.

.

INTRODUCTION

Despite the multi ethnic nature of Nigeria and its social interactions of cultures and migration, the country still remains one of the few African nations with strong values attached to traditions. The three main ethnic groups namely, Hausa, Igbo and Yoruba as highlighted by [1]–[3], can in most cases be distinguished in terms of the attires they use and the style of their dressing or more generally in their physical appearances. The categorization of these ethnic groups based on the above factors is fast changing due to socialization, migration and intermarriages. Therefore, with time these characteristics may drastically be reduced due to significant acceptance of other's way of life. But another physiological characteristic of humans is found in their sound production [4] which contains interesting features that can be used to classify an individual ethnicity. For example, the accent of a speaker or people from a certain geographical location is distinct.

This intrinsic parameter of speech in the first language will always persists in the second language due the strong doggedness to one's mother's tongue (MT) [3]. Thus, MT is most noticeable and long-lasting parameters that one can always finds in the area of pronunciation, as compared to grammar and vocabulary. A lot of research findings support this view such as [3], [5] – [7]. This can only change if a person spends more time outside his/her native environment than in his environment of origin [8], [9]. The inherent accent of a person can be useful in many application areas of speech processing; such as speaker identification, speech recognition, speech to text recognition, etc. To study the differences in accent among many tribal groups in Nigeria especially for the purpose of automatic speech analysis, an audio database is required that will contains all the tribal variations of interest. However, the literature search conducted has proved that such database is not available to research community. The only serious and biggest

Corresponding author: Rabiu. H. ✉ rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

corpus available was developed to test the performance of a speech recognition system based on American English [10].

Nigeria with over 180 million people from the census of 2013, and also having over 400 languages as highlighted by [2], [3] of which three main languages are considered as Nigeria's major languages HIY certainly need to have its own corpus for the purpose of research. In Nigeria, individuals from these three (3) ethnic groups have their specific ways of pronouncing English words which is unique and is generally referred to as Nigerian English [11] – [13]. Therefore, this solitary of English words pronunciation across the three (3) tribal groups could lead to development of algorithms that can automatically classify the speaker's ethnic group based on the words pronounced. The lack of algorithm developed to discriminate and improve the performance of speech recognition systems in Africa and Nigeria in particular is largely due unavailability of local Corpus that takes into account the varieties of accents within a country. The goal of this study is therefore to develop an isolated words database from the three main tribal groups of Nigeria for research and speech processing purposes. The remaining of this paper is structured as follows: Section 2 presents the methodology used, while section 3 presents the Analysis of the results obtained. Finally, Conclusion is given in Section 4.

METHODOLOGY

Database Development

The use of short recorded words has been demonstrated on speaker accent identification in many researches [14] – [19]. To identify and select the proper words for this study, the services of specialist in the linguistic domain was invoked to guide in the selection of English words that are pronounced differently across the three ethnic groups. In developing this database, we adopted the method used by [20] – [23] in developing similar works for different languages and ethnic group across the world. Figure 1 illustrate the general frame work of the method. The database was developed from 72 speakers selected from the three main ethnic groups, where each speaker was asked to recite words in a reading style. The audio signals were recorded using a personal computer with Audacity software having a sampling rate of 16 kHz and sample size of 16 bits.

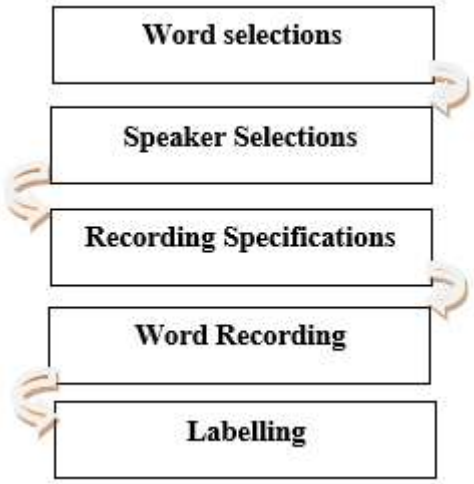

Figure1: Database Development Methodology

Word Selections

The words selection was guided by the expert from Nigerian languages Department of Federal College of Education Kano. The English words selected for this work were words whose phonemes have been widely studied by many researchers [5], [8], [9], [12], [20] in the field of linguistic. Results of such studies, have shown that, there exists a strong correlation between the words and pronunciation pattern by each ethnic group involved.

Yoruba English Words Pronunciations

Due to the absence of the sound of letter /h/ in Yoruba phonetic structure, word such as hospital which starts with the letter /h/, the Yoruba speak simply omitted /h/ while pronouncing the word hospital [3], [20]. Therefore, words such as human, house, high, hardly, hospital, hand, help, heat were included in the database. Furthermore, Yoruba native speakers of English tend to pronounce /ch/ sound as /s/, due to the absence of /sh/ in their phonetic structure. Then words such as child, and champion are pronounced by a Yoruba speaker as /Shampion/ and /Saild/ respectively were added to the corpus. In addition, due to the absence of the sound of letter /z/ in Yoruba, a Yoruba speaker replaces it by its closest in sound that is letter /s/. Thus, he

Corresponding author: Rabiu. H. ✉ rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

pronounces the word zoo as /soo/. Other phonemes that are absence in Yoruba ethnic group are the nasalized English vowels that are preceded by nasal consonants [12]. For instant, they pronounce the word /morning/ as /morin/ due to the absence of velar nasal voiced /g /. Similarly, they pronounce the word very as /feri/.

Hausa English Words Pronunciations

Experiments by [3], [20], revealed that native speakers of Hausa tend to replace bilabial voiceless /p/ with labiodentals (sounds produced by both the teeth & lips) fricatives voiceless /f/ and vice versa, for example the words pencil, food, four. Free, problem, primary, put are pronounced /fencil/, /pood/, /pour/, /pree/, /froblem/, /frimary/and /fut/ respectively. Therefore, such words among many others are included for this experiment. The Hausa speakers are also prone to interchanging the bilabial voiced /b/ and the labiodentals /v/ in words such as visitor, briefly, better, boat, etc which they pronounce as follows: bisitor, vriefly, vetter, voat etc. This Interchange of /v/ and /b/ sounds led to their selections as part of this experiment. In the same way, due to the absence of labiodentals sound /th/ in Hausa, the Hausa speakers of English substitute it with close sound. For example, the word /this/ is pronounced/zis/. The /th/ sound is therefore pronounced /z/.

Igbo English Words Pronunciations

On the other hand, Igbo speakers of English, even among some well-educated ones, tend to transfer vowel system of their native language into English [12], [13]. The Igbo speakers have what is called 'light' and heavy vowels''. The heavy vowels are stressed upon anywhere they appear. These are the five pure English vowels (a, u, i, o, e). For example, words such as school, food, morning, etc are dragged wherever those vowels appear. Also, in Igbo phonetic structure, a consonant cannot stand alone; each consonant must be followed by a vowel. Therefore, using the above results analysis, a total of 36 words were selected as presented in Table1 in which each ethnic group is accorded 12 words.

Table1. Selected words

SN	HAUSA	IGBO	YORUBA
1	Human	Food	Busy
2	Child	School	This
3	Zoo	Good	Visitor
4	House	Business	Very
5	Champion	Taxation	That
6	High	Maximum	Four
7	Hardly	Pencil	Free
8	Hospital	Morning	Government
9	Hand	Thank	Problem
10	Help	Think	Primary
11	Shout	Evening	Put
12	Heat	Wrapper	Biefly

PARTICIPANTS SELECTION

A questionnaire was formulated as regarding the objective of the study (to design and develop a system that can be able to distinguish the three ethnic groups), we stratified the population into three strata (gender, age and educational qualifications). More than 200 questionnaires were administered in order to achieve a good sampling of a speaker population, the questionnaires also attempt to capture participant biodata. In addition, it also records the speakers' background with regard to his exposure to speech processing systems. Therefore, several questionnaires were rendered as shown in Table 2 Generally, the participants are students from two higher institutions located in Kano/Nigeria. The institutions are: Bayero University Kano (BUK) and Federal College of Education, Kano (FCE) that cut across twenty-five (25) States of Nigeria. The participant state distribution is as depicted in Figure 2, with Kano, Kwara and Imo States having the highest number of participants 25%. 16% and 10%. Respectively.

Corresponding author: Rabiu. H. rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

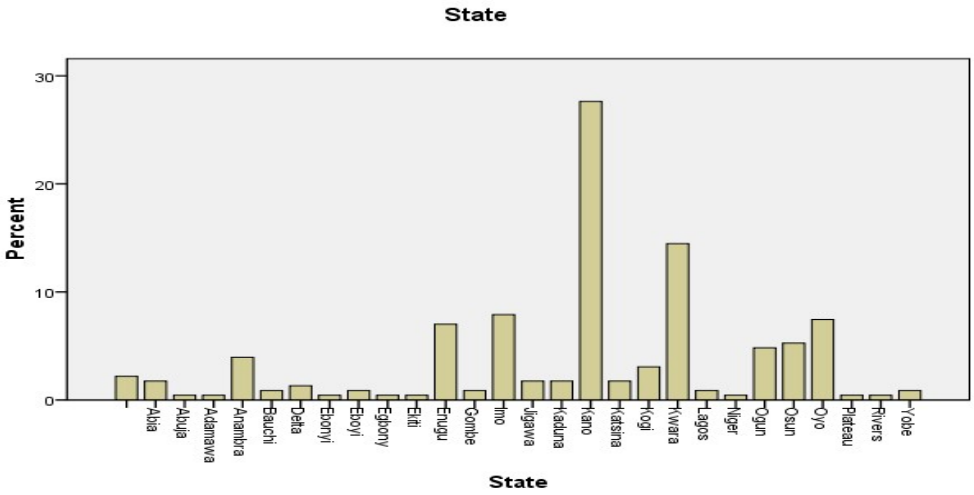

Figure 2: Participants' vs States Distribution

To make the proposed database more robust and gender balance. We made male female participation almost equal as illustrated in Table 2. This is to account for the serious variations in voice between male and female, noting that the mean fundamental frequency, which is associated with the perceptual notion of pitch, is commonly considered as the major difference between adult male and female voices. Mean F0 would be around 120 Hz for men and 200 Hz for women [21] – [23] and also for inter-speaker variation as in [24].

Table 2. Male Female participation ratio

		Frequency	Percent
Valid	Female	113	49.6
	Male	115	50.4
	Total	228	100.0

RECORDING SPECIFICATIONS

The speeches were recorded at the Bayero University audio visual studio, which is built using sound proof materials. The recording studio was divided into two parts; with the inner part used as the recording room. The recording was done using digital recording equipment; a desktop computer with audacity software installed was used in recording the speeches. The system is calibrated to record the sound at 16 MHz with sampling rate of 16bits A condenser microphone was used as the main transducer that convert the sound energy into electrical signal. The microphone is placed at 10 to 15cm away from the speaker, detailed system specification is given in Table 3. The recorded data is directly stored on the computer's memory. Participants were asked to read the isolated words (36 words) with each word repeated five times.

We made two recording sessions twice in a day to suit the participants schedules. The recording was done either in the mornings or afternoon at the participant's discretion and convenience. Both sessions were recorded under a temperature ranging from 22 to 30-degree census. The recordings lasted for about 20 minutes per participant. The participants were recorded as they come one after the other at different time and days. A little time was given between each word in order to allow the participants to rest. Table3. Presents the recording parameter

Table 3. General system specifications

.SN	DESIGNATION	VALUE
1	Desktop	Pentium R at 240 GHZ
2	Microphone	Ultra gain mIC 200Behringer
3	Studio speakers	

Corresponding author: Rabiu. H. rabiuhabib@gmail.com *Department of Electrical Engineering, Bayero University, Kano.*

	Multiplexer	Behringer 24 bit. Multi-FX processor
4	Sampling rate	16 bits
5	Sampling frequency	16000
6	Distance from MIC	10-15 cm

Figure 3: Audio recording at the Oral Documentation Unit (BUK)

DATA RECORDING

The 46 selected isolate words were read by 72 participants, the words were shown to the speaker from a transparent glass outside the recording room. Figure 3 illustrates the recording session at the oral documentation unit of the Department of Nigerian Language, Bayero University Kano (BUK). For each word shown to the participant, he or she repeats that same word five times at different intervals to avoid mixing up or change of intonation. This way, all participants read the words which sum up to a total of 12,960 audio samples recorder.

SOFTWARE USED

Audacity software v2.3.3 was used for the recording, the lab technician configured the software settings to appropriate levels. A participant was recorded on new file which is exported and saved using" .wav" file format with the name of the participant on it. The software environment is as shown in Figure 4.

Corresponding author: Rabiu. H. ✉ *rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.*

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 10(4), DECEMBER, 2022
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com

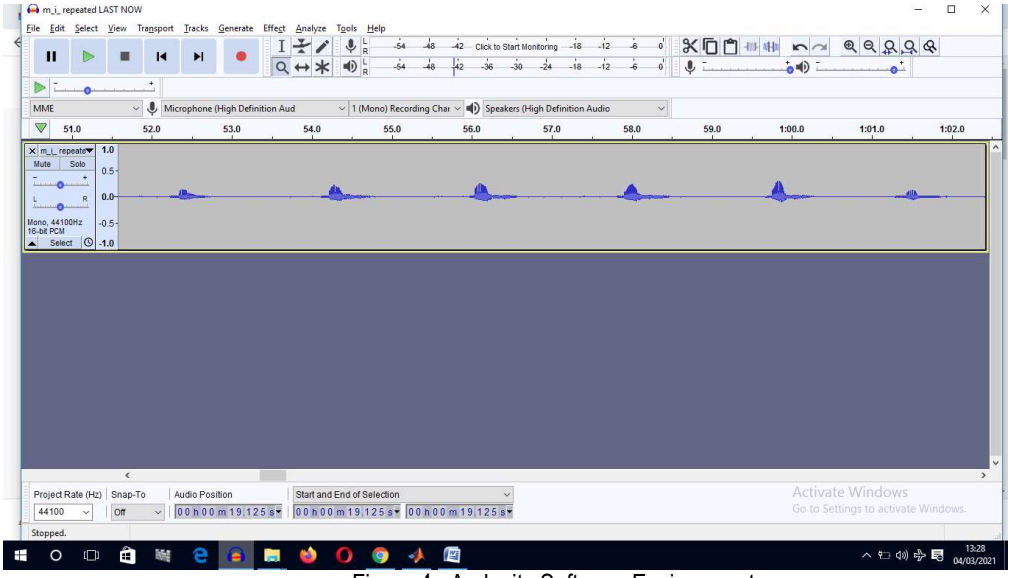

Figure 4. Audacity Software Environment

LABELLING

As describe by [25] the process of aligning a symbolic description of some speech item to the recorded signal itself is to help in distinguishing labelling of one speaker from the other. The method used by [24] is adopted in this work as shown in Figure 5. In this scheme each recorded piece is given a label that include the speaker's identity. For example: M-H-1 stand for (M) male, (H) Hausa, and (1) for first sample in Hausa group. or F-H-5 to mean female Hausa sample number 5. Similarly, F-Y-3, M-I-9 are to mean female Yoruba sample 3 and male Igbo sample 9 respectively. Figure 5 and 6 give a more detail on the labelling method used.

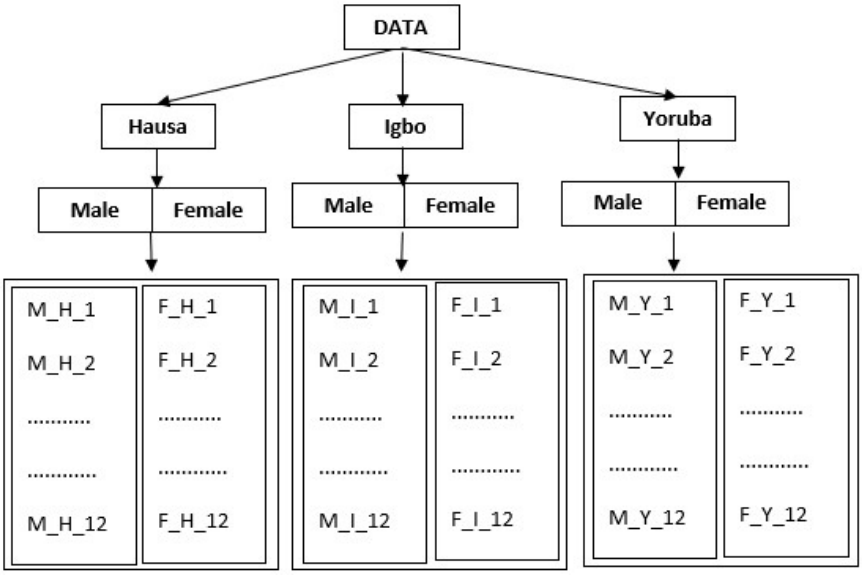

Figure 5: Organizational Labelling of the Recordings

Corresponding author: Rabiu. H. rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

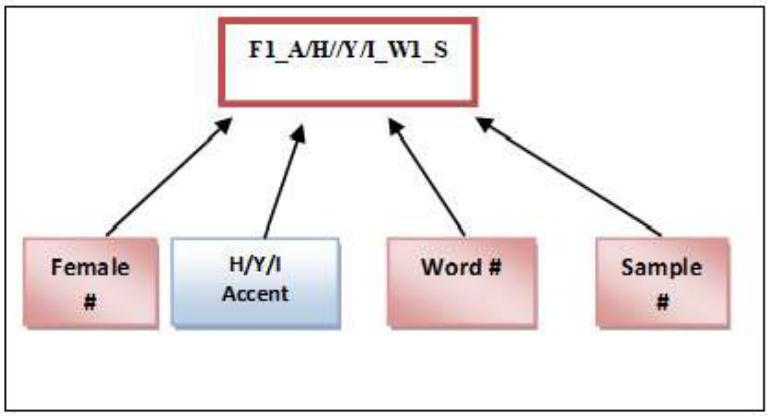

Figure 6: Labelling Template

DATABASE SUMMARY

At the end of recording, labelling and validation of the recorded samples, some samples that were found to be of poor quality were discarded. Finally, we are left with 11,472 audio samples from 72 participants as summarised in Table 4.

Table 4: Database. Summary

Accents	**N0. Speakers**	**Type of utterances**	**N0. Utterances**	**N0. Words**	Recording specifications	
Hausa Group	12 males & 12 females	Isolated word	4320	12	Sampling freq 16 Khz	Sampling Rate 16 Bps
Igbo Group	12 males & 12 females	Isolated word	4320	12		
Yoruba Group	12 males & 12 females	Isolated word	4320	12		
TOTAL	**72**	**Isolated word**	**12 960**	**36**		

POTENTIAL APPLICATION AREAS OF THE DATABASE

The developed database can be used in the following application areas: -

1. Can be used for biometric researches.
2. Can help the security agencies to narrow down the tracking of a suspect.
3. Can serve as an identity stance on issues in court.
4. A support telephony-based help Systems
5. Can be applied in crime investigation to determine the ethnicity of a suspect.
6. A step to improving the performance of ASR systems by adapting the ASR with accented Nigerian English.
7. Can be used as a pédagogique tool to improve second language learning.
8. In addition to ASR applications, accent identification is also useful for forensic speaker profiling by identifying the speaker's regional origin and ethnicity in applications involving targeted marketing.

CONCLUSION

In this work, we developed a marked accent database constituting the speeches of three main tribal groups of Nigeria, whose pronunciation patterns can easily be used in identifying their ethnicity. We identified particular phonemes that are unique to each of the participating ethnic group and selected words in

Corresponding author: Rabiu. H. rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

which those phonemes can clearly be seen as they are pronounced. At end of the recordings 12,960 audio samples were recorded which sufficiently captured the ethnic variations of the 3 languages involved. The developed databased will open new paradigms of research in the areas of speech analysis, sound based biometric studies, provide tool for ASR performance improvement and can as well serves as pedagogue tool to improve second language learning. In general, the work will give research community another opportunity to explore and discover newer approaches to speech recognition and analysis. Future work should focus on developing a continuous speech database from a larger population sample that could be used for more challenging problem.

ACKNOWLEDGMENT

We would like to acknowledge the tremendous support and cooperation given to us by Prof. Saidu Muhammad Gusau and the Department of Nigerian languages BUK for availing their ultra-modern oral documentation studio for the recordings.

REFERENCE

[1] U. Cunningham, "Using Nigerian English in an international academic setting," Res. Lang., vol. 10, no. 2, pp. 143–158, 2012.

[2] A. R. Mustapha, Ethnic structure, inequality and governance of the public sector in Nigeria. United Nations Research Institute for Social Development Geneva, Switzerland, 2006.

[3] J. M. Patrick, M. E. E. Education, and M. Sui, "Mother-tongue interference on English language pronunciation of senior primary school pupils in Nigeria: Implications for Pedagogy," Lang. India, vol. 13, no. 8, pp. 281–298, 2013.

[4] T. R. Agus, S. J. Thorpe, C. Suied, and D. Pressnitzer, "Characteristics of human voice processing," in Proceedings of 2010 IEEE International Symposium on Circuits and Systems, 2010, pp. 509–512.

[5] P. Kiatkheeree, "Learning environment for second language acquisition: through the eyes of English teachers in Thailand," Int. J. Inf. Educ. Technol., vol. 8, no. 5, pp. 391–395, 2018.

[6] S. S. Olanipekun, D. Atteh, J. A. Zaku, and P. E. Sarki, "Mother tongue and students' academic performance in English language among secondary school students," Int. J. Lang. Lit. Cult., vol. 1, no. 1, pp. 1–6, 2014.

[7] T. U. Sa'ad and R. Usman, "The causes of poor performance in English language among senior secondary school students in Dutse Metropolis of Jigawa State, Nigeria," IOSR J. Res. Method Educ., vol. 4, no. 5, pp. 41–47, 2014.

[8] M. O. M. Chitulu and Q. U. Njemanze, "Poor English pronunciation among Nigerian ESL students; the ICT solution," Int. J. Lang. Lit., vol. 3, no. 1, pp. 169–179, 2015.

[9] C. U. Omego, "Influence of Environment on a Child's Acquisition of English as a Second Language and the Gradual Extinction of Nigerian Languages: A Study of Children of selected schools in Choba, Nigeria," AFRREV IJAH Int. J. Arts Humanit., vol. 3, no. 3, pp. 146–161, 2014.

[10] S. A. Amuda, H. Boril, A. Sangwan, J. H. Hansen, and T. S. Ibiyemi, "Engineering analysis and recognition of Nigerian English: an insight into low resource languages," Trans. Mach. Learn. Artif. Intell., vol. 2, no. 4, pp. 115–28, 2014.

[11] F. Foluke, "Perceptual convergence as an index of the intelligibility and acceptability of three Nigerian English accents," Int. J. Appl. Linguist. Engl. Lit., vol. 1, no. 5, pp. 100–115, 2012.

[12] O. R. Modupeola, "Varieties of English Language in Nigeria: A Tool for Character Stratification in Wole Soyinka's The Beatification of Area Boy.," IOSR J. Humanit. Soc. Sci. IOSR-JHSS, vol. 19, no. 4, pp. 86–89, 2014.

Corresponding author: Rabiu. H. ✉ rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.

[13] O. K. Olaniyi and U. E. Josiah, "Nigerian accents of English in the context of world Englishes," World J. Engl. Lang., vol. 3, no. 1, p. 38, 2013.
[14] F. Biadsy, Automatic dialect and accent recognition and its application to speech recognition. Columbia University, 2011.
[15] Z. Ge, "Improved accent classification combining phonetic vowels with acoustic features," in 2015 8th International Congress on Image and Signal Processing (CISP), 2015, pp. 1204–1209.
[16] L. W. Kat and P. Fung, "Fast accent identification and accented speech recognition," in 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), 1999, vol. 1, pp. 221–224.
[17] A. Phidd, P. Thimmappa, R. Sauther, and S. Vijayakumar, "Speech Database/Tool System and Preliminary Accent Recognition Study".
[18] D. Stantic and J. Jo, "Accent identification by clustering and scoring formants," Int. J. Comput. Syst. Eng., vol. 6, no. 3, pp. 379–384, 2012.
[19] C. Vimala and V. Radha, "Isolated speech recognition system for Tamil language using statistical pattern matching and machine learning techniques," J. Eng. Sci. Technol. JESTEC, vol. 10, no. 5, pp. 617–632, 2015.
[20] U. Josiah, H. Bodunde, and E. Robert, "Patterns of English pronunciation among Nigerian university undergraduates: Challenges and prospects," Int. J. Bus. Humanit. Technol., vol. 2, no. 6, pp. 109–117, 2012.
[21] L. C. Oliveira, S. Paulo, L. Figueira, C. Mendes, A. Nunes, and J. Godinho, "Methodologies for Designing and Recording Speech Databases for Corpus Based Synthesis.," 2008.
[22] A. Pawi, "Modelling and extraction of fundamental frequency in speech signals," PhD Thesis, Brunel University School of Engineering and Design PhD Theses, 2014.
[23] E. Pépiot, "Voice, speech and gender:. male-female acoustic differences and cross-language variation in english and french speakers," Corela Cogn. Représentation Lang., no. HS-16, 2015.
[24] H. Melin, "Databases for speaker recognition: Activities in COST250 working group 2," COST 250-Speak. Recognit. Teleph. Final Rep. 1999, 1999.
[25] J. Hennebert, H. Melin, D. Petrovska, and D. Genoud, "POLYCOST: a telephone-speech database for speaker recognition," Speech Commun., vol. 31, no. 2–3, pp. 265–270, 2000.

Corresponding author: Rabiu. H. ✉ rabiuhabib@gmail.com Department of Electrical Engineering, Bayero University, Kano.