Ebira English in Nigerian Supersystems: Inventory and Variation

Dissertation

zur Erlangung des akademischen Grades doctor philosophiae (Dr. phil.)

vorgelegt der Philosophischen Fakultät der Technischen Universität Chemnitz


Chemnitz, 30.01.2017
Dean: Prof. Dr. Stefan Garsztecki

Supervisor: Prof. Dr. Josef Schmied

External Examiner: Prof. Dr. Albertus J van Rooy
To my father and mother
# Content

List of Figures iv
List of Tables vi

1 **Nigerian English Varieties: Conflicts and Emergence** ......................... 1
   1.1 The Ebira English Subsystem .................................................. 3

2 **Through Evolution, Diversity and Powers of English in Nigeria** ............. 8
   2.1 Functional Powers of English since Contact ................................ 11
   2.2 The Ebira People and English in Ebiraland .................................. 18
   2.3 The Benue Congo Phylum .......................................................... 20
   2.4 Ebira and Yoruba Vowel Systems .............................................. 22
   2.5 NigE Varieties: Ethno-linguistic Nexus ....................................... 24
      2.5.1 Level of Education ............................................................ 26
      2.5.2 The Lects and Quest for Standardisation ............................... 29
   2.6 Overview of NigE Vocalic Inventory: Monophthongs .......................... 33
      2.6.1 High, Low Back and Central Vowels .................................... 37
      2.6.2 COMMA/LETTER Lowering .................................................. 45
      2.6.3 NURSE Lowering, Backing or Fronting .................................. 47
   2.7 Diphthongs ................................................................................. 48
   2.8 Research Questions ...................................................................... 50

3 **Research Design** .............................................................................. 52
   3.1 Speakers ....................................................................................... 52
   3.2 Sampling ....................................................................................... 53
      3.2.1 Age ....................................................................................... 56
      3.2.2 Gender ................................................................................... 58
      3.2.3 Education .............................................................................. 59
   3.3 Sociolinguistics Questionnaires and Interviews .................................. 60
3.3.1 The Word List Citation .......................................................... 62
3.3.2 The Reading Passage ............................................................ 64
3.3.3 The Sociolinguistic Interview ................................................. 67
3.4 Recording .................................................................................. 70
  3.4.1 Segmentation and Measurements ........................................... 70
  3.4.2 Formants Measurement ....................................................... 72
  3.4.3 Data Cleaning ......................................................................... 76

4 Theoretical Frameworks ................................................................ 81
  4.1 Introduction .............................................................................. 81
    4.1.1 Mufwene’s Feature Pool .................................................. 82
    4.1.2 Linguistic Species and Accommodation Theory ................. 85
    4.1.3 Sociolinguistic Identity and Accommodation .................... 87
    4.1.4 Dynamic Model ............................................................... 89
    4.1.5 NigE in Schneider’s PCE .................................................. 92
    4.1.6 Conclusion ......................................................................... 96
  4.2 Data Normalisation ................................................................... 97
    4.2.1 Statistical Assumptions and Modelling ............................... 103
    4.2.2 Linear Regression Models ............................................... 104
      4.2.2.1 Choosing Predictors for Statistical Modelling .............. 110
  4.3 Methods for ‘Mergers’ ............................................................... 116
    4.3.1 Mixed Effects Modelling .................................................. 118
    4.3.2 Random Intercept in Mixed Models and By-speaker Analyses ... 120
    4.3.3 Rbrul for Mixed Effects Regression .................................. 121
    4.3.4 Interpreting Rbrul Outputs .............................................. 123

5 Analysis of EEng Monophthongs .................................................... 126
  5.1 Overview ................................................................................. 126
  5.2 Model Fitting and Linear Regression Analysis ............................ 103
    5.2.1 High Vowels: KIT & FLEECE ....................................... 135
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.2</td>
<td>GOOSE and FOOT</td>
<td>140</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Low Back Vowels: LOT, THOUGHT AND STRUT</td>
<td>146</td>
</tr>
<tr>
<td>5.2.4</td>
<td>STRUT and Low Back Vowels</td>
<td>151</td>
</tr>
<tr>
<td>5.2.5</td>
<td>Low Central Vowels</td>
<td>154</td>
</tr>
<tr>
<td>5.2.5.1</td>
<td>Analysis of TRAP, BATH and lettER</td>
<td>157</td>
</tr>
<tr>
<td>5.2.5.2</td>
<td>The Status of lettER and BATH</td>
<td>161</td>
</tr>
<tr>
<td>5.2.6</td>
<td>NURSE</td>
<td>164</td>
</tr>
<tr>
<td>5.2.7</td>
<td>FACE, GOAT and CURE</td>
<td>167</td>
</tr>
<tr>
<td>5.3</td>
<td>Monophthongal Inventory in Ebira English System</td>
<td>169</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Extralinguistic Variables</td>
<td>173</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Regression Analysis</td>
<td>175</td>
</tr>
<tr>
<td>5.3.2.1</td>
<td>KIT and FLEECE</td>
<td>176</td>
</tr>
<tr>
<td>5.3.2.2</td>
<td>GOOSE and FOOT</td>
<td>179</td>
</tr>
<tr>
<td>5.3.2.3</td>
<td>USE</td>
<td>180</td>
</tr>
<tr>
<td>5.3.2.4</td>
<td>THOUGHT and STRUT</td>
<td>181</td>
</tr>
<tr>
<td>5.3.2.5</td>
<td>BATH and lettER</td>
<td>183</td>
</tr>
<tr>
<td>5.3.2.6</td>
<td>NURSE</td>
<td>185</td>
</tr>
<tr>
<td>5.3.2.7</td>
<td>DRESS</td>
<td>187</td>
</tr>
<tr>
<td>5.4</td>
<td>Social Differentiation in EEng Vowels</td>
<td>188</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion</td>
<td>191</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary</td>
<td>192</td>
</tr>
<tr>
<td>6.2</td>
<td>Limitations</td>
<td>196</td>
</tr>
<tr>
<td>6.3</td>
<td>Outlook</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td><strong>Bibliography</strong></td>
<td>199</td>
</tr>
<tr>
<td></td>
<td><strong>Deutsche Zusammenfassung</strong></td>
<td>219</td>
</tr>
<tr>
<td></td>
<td><strong>Appendix</strong> (on CD)</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1: An index map for some Nigerian languages .......................................................... 21
Figure 2.2: Olaniyi’s pyramid for NigE varieties (Olaniyi 2014:237) ........................................ 26
Figure 2.3: An illustrative lects ladder for I am eating in Jamaican speech continuum .......... 29
Figure 2.4: Difference in duration between stressed –ne– of phonetics and the unstressed –ne– in phonetician .......................................................... 46
Figure 3.1: Style shifting of (oh) by male and female speakers across three socioeconomic groups .......................................................................................... 58
Figure 3.2: r-realisation in New York City by socioeconomic classes and speech styles........... 64
Figure 3.3: F1 & F2 formant features of natural and synthetic vowels. The comet plots labelled /el/, /il/ and /sl/ represent the mean formant onsets and offsets of the vowels as produced in isolated-word /bVp/- context in Western Canadian English ...................... 73
Figure 3.4: Textgrid showing monophthongal realisation or absence of glide in GOAT (in informal speech (MUO1)) & FACE (in reading passage (FUY2)) .............................................. 76
Figure 3.5: A spectra cut of a speaker at the beginning paragraph of the The Boy Who .......... Cried Wolf .................................................................................................................. 77
Figure 3.6: F1 & F2_normalised plot for all monophthongs prior to final cleaning and removal of tokens of less than 70ms and above 220 ms from the dataset.............. 79
Figure 4.1: A model based on Mufwene’s notion of language ecology and evolution ............ 87
Figure 4.2: Idealised vowel triangle for the construction of the centroid S. i=min F1, .......... max F2; a=max F1; u’=min F1, min F2, where F1 (u’) and F2 (u’) =F1 (i) ...................... 99
Figure 4.3: F1/F2 residual plots of FLEECE&KIT showing mild violation of linearity assumption ............................................................................................................ 105
Figure 4.4: Normal Q-Q plot for normalised F1 of FLEECE&KIT ......................................... 108
Figure 4.5: Normal Q-Q plot for F1 normalised values of KIT .................................................. 108
Figure 4.6: Decision Trees of predictors for FLEECE/KIT_F1 showing the initial splits for gender as most influential ................................................................. 113
Figure 4.7: VarImp (Variable importance) plots for FLEECE/KIT_F1 with the highest prediction for gender ......................................................................................... 115
Figure 5.1: Normalised F1/F2 plot of EEng vowel system (including diphthongs) for all speakers ............................................................................................................ 127
Figure 5.2: F1/F2 plot of the historically merged sets in NigE................................. 129
Figure 5.3: Scatter plot showing USE_GOOSE distribution in the F1&F2 ......................... 141
Figure 5.4: All speakers plot of FOOT & GOOSE tokens separated by preceding and following phonological co-texts ............................................................. 143
Figure 5.5: F1 & F2 contour plot of LOT and THOUGHT in the following and preceding co-texts .................................................................................................................. 149
Figure 5.6: Contour plots showing marginal differentiation between STRUT & THOUGHT sets (more in F1) by style = formal and informal speech styles following co-texts.............. 154
Figure 5.7: F1/F2 plot showing similar pattern of distribution for TRAP & BATH in both dimensions .................................................................................................................. 157
Figure 5.8: F1/F2 plot of TRAP and BATH separated by gender, showing a close proximity between the two phonemes .......................................................................................... 161
Figure 5.9: Plots of lettER and BATH vowels in the following co-texts indicating some overlap and a differentiation in F1 ................................................................. 163
Figure 5.10: NURSE plot for realisation patterns based on orthographical taxonomy ........... 166
Figure 5.11: F1/F2 plot of glide trajectory for complex vowels in EEng – showing comparatively weaker glides for FACE and GOAT ............................................. 169
Figure 5.12: Visual summary of monophthongal inventory in Ebira English vowels. Both FACE and GOAT have diphthongal status in the analysis .................................................. 170
Figure 5.13: Graphical illustration of gradient differences in KIT realisation between gender speakers and age_groups ................................................................. 177
Figure 5.14: Plot showing the trajectories of differentiations in BATH and lettER – by gender and style ........................................................................................................... 185
LIST OF TABLES

Table 2.1: Language distribution in Nigerian states................................................................. 14
Table 2.2: Interview excerpt on the attitude and spread of English among Nigerians......... 17
Table 2.3: Ebira vowel harmony sets .................................................................................... 23
Table 2.4: Ethnodemographic patterns in Nigeria.................................................................... 25
Table 2.5: Ugorji’s (2010) summary of Nigerian Received Pronunciation ......................... 31
Table 2.6: Summary of ‘principal vowel errors’ in NigE phonemes ..................................... 34
Table 2.7: Monophthongal inventory of Educated Hausa (Northern) and Yoruba/Igbo (Southern) English ....................................................................................................... 36
Table 2.8: A list of context independent and dependent merger paradigms in NigE .......... 39
Table 2.9: Vowel realisations in educated Nigerian English (B) and other possible variants .. 42
Table 2.10: Perceptual frequency of schwa in comma /ə/ in unstressed syllables ............... 46
Table 2.12: Diphthong realisations in educated Nigerian English (B) and other possible variants .................................................................................................................................... 49
Table 2.13: Patterns of diphthong reduction in African Englishes ........................................ 50
Table 3.1: Speaker name, gender, age-category, education degree, other Nigerian languages, spoken, chronological age, age of exposure to English and years of exposure ..................... 55
Table 3.2: Distribution of respondents according to age and sex ......................................... 56
Table 3.3: Trend of linguistic change in the individual and the community ......................... 57
Table 3.4: Token distributions according to lexical sets from The Boy who Cried Wolf ...... 66
Table 4.1: The applicability of constituent components drawn from the Dynamic Model in Expanding Circle countries and emergent contexts ............................................................................ 94
Table 4.2: Estimation of collinearity among predictors for FLEECE/KIT in F1 & F2 ........ 111
Table 5.1: Phonological classifications in preceding and following co-texts ......................... 132
Table 5.2: Initial cross-tabulation of all FLEECE & KIT tokens with preceding phonological contexts ........................................................................................................................................ 133
Table 5.3: Initial cross-tabulation of all FLEECE & KIT tokens with following phonological contexts .......................................................................................................................... 133
Table 5.4: Adjusted distribution both co-texts (complex model).......................................... 134
Table 5.5: F1 regression results for KIT & FLEECE in following co-texts and preceding co-texts .......................................................................................................................... 137
Table 5.6: F1 results for KIT & FLEECE for the complex model ........................................ 139
Table 5.7: Initial cross-tabulation of all GOOSE & FOOT tokens with preceding phonological contexts before adjustment .................................................................................................................. 141
Table 5.8: Significant predictors for FOOT & GOOSE in F2 for preceding co-texts and complex model .......................................................................................................................................................................................................................................................... 144
Table 5.9: LOT & THOUGHT in F2 for preceding co-texts and complex model ........... 150
Table 5.10: STRUT & THOUGHT results in F2 for complex model ......................... 152
Table 5.11: Results of linear model assumptions for a MANOVA regarding the distribution of gender and mean formant values of TRAP and BATH on four degrees of freedom.... 159
Table 5.12: Post-adjustment distribution of NURSE tokens according to following co-texts. 165
Table 5.13: Significant predictors in KIT realisation ............................................................ 177
Table 5.14: Results of significant variables in the analysis of THOUGHT and STRUT........ 182
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NigE</td>
<td>Nigerian English</td>
</tr>
<tr>
<td>EEng</td>
<td>Ebira English</td>
</tr>
<tr>
<td>n.d.</td>
<td>no date</td>
</tr>
<tr>
<td>GCE</td>
<td>General Certificate Examination</td>
</tr>
<tr>
<td>RP</td>
<td>Received Pronunciation</td>
</tr>
<tr>
<td>SBE</td>
<td>Southern British English</td>
</tr>
<tr>
<td>ESL</td>
<td>English as Second Language</td>
</tr>
<tr>
<td>VIF</td>
<td>Variable Inflation Factor</td>
</tr>
<tr>
<td>MC</td>
<td>Middle Class</td>
</tr>
<tr>
<td>LMC</td>
<td>Lower Middle Class</td>
</tr>
<tr>
<td>UP</td>
<td>Upper Class</td>
</tr>
<tr>
<td>F1</td>
<td>Formant one</td>
</tr>
<tr>
<td>F2</td>
<td>Formant two</td>
</tr>
<tr>
<td>ED</td>
<td>Euclidean Distance</td>
</tr>
</tbody>
</table>
1 Nigerian English Varieties: Conflicts and Emergence

Perhaps the most controversial issue in English Language Studies in Nigeria is that of Nigerian English. Scholars such as Banjo, Adetugbo, Adesanoye and Odumuh affirm the existence of Nigerian English...Nigerian English has developed distinct phonetic, phonological, lexical and syntactic characteristic which are quite stable and which cannot be regarded as deviation from a native norm which Nigerians do not, in any case, aspire to approximate (Jibril 1987 cited in Schmied 1991a:175).

Scholars have, over the years, shown vigorous commitments to the analysis of what is now acknowledged as Nigerian English (hence NigE). This concession, of course, is a by-product of theoretical conflicts, some of which are obvious till date. The split has been between language experts who in their works have either conceded or dismissed the concept of Nigerian English system, thus forming the ‘right’ and ‘left’ of the praxis (Adegbite 2010:14). Generally, the ‘right’ comprises followers of Prator (1968), who often demand the exclusion of NigE from the ranks of ‘standard’ varieties\(^1\). Notable voices from this camp include Salami (1968) and Vincent (1974) who brand the variant as nothing other than bad English or mere reckless errors of usage. Oji (n.d.) cited in Jibril (1987: 47), is also reported to have denounced it, declaring that ‘the death knell of Nigerian English should be sounded loud and clear, as it has never existed, and does not now, and will never see the light of day’.

As far as NigE is concerned, major foundations for prescriptivism straddle the works of Nuttall (1961&1965) on Hausa English, Schafer (1967), Afolayan (1968) on Yoruba English, and Tiffens (1974) on Hausa and Yoruba accents. Their views, during these periods, were likewise encouraged by broad efforts of para-linguistic agencies involved with structuring and training learners on how the ‘proper’ English ought to be spoken – part of which is reflected in an assessment report of 1961 General Certificate Examination (GCE).

\(^1\) In divergent movements notoriously dubbed as The Pratorian Guards and the Liberalist Linguists, two hotly polarised camps have discoursed on the elements that constitute the structure and status of the English Language outside its native home. Taking a rather fanatical and prescriptive stance is Prator (1968), who in his publication, *The British Heresy in TESOL*, condemns the unmasterly handling of English language among non-native speakers attempting to domesticate it around the world. He plainly calls for adherence to the ‘pure’ British English model among native and non-native speakers alike. Of the Liberalists is Kachru (1986) who in his article, *Models of English for the Third World: White Man’s Linguistic Burden or Language Pragmatics?*, accused Prator of the famous nine ‘sins’, which include: his wrong perception of English in its global robes, ‘ethnocentrism’, clear disregard for linguistic interference and language dynamics, other cline of Englishes, and so on.
For example, an assessor’s comments on performance in Oral English reads: ‘…bad pronunciation led to such spellings as order (other), efford (effort), all works (walks) of life, a match – (march-past), leave (live), modern (morden), reach (rich), prices (prizes), and communnal/communial (communal)’ (University of London GCE Report 1961:33). Also, pedantic tags such as mispronunciation, incorrect placement and wrong realisation of English pronunciation among Nigerian speakers appear on almost every section of Tiffen’s (1974) perception analysis of NigE systems. Another veteran of the Pratorian Guards is Fakoya (2004) who, in his fierce critique of typical lapses in NigE, dismisses them as mediolects – meaning the lects of mediocre speakers. Other studies in this mould include series of contrastive investigation of NigE against established varieties such as the Received Pronunciation (RP) or the Southern British English (SBE).

In opposition, those of the ‘left’ do not only confirm that NigE exists, but have also made modest attempts at defining its distinctive features. In so doing, they foreground the linguistic properties that tell apart the Nigerian systems from others, especially from the so-called native structures. Further aspirations of scholars on this divide have likewise motivated the continuing drive towards standardising, and possibly, stabilising the variety (see Adetugbo 1977, Awonusi 1990, Odumuh 1987 and Schneider 2004). Very basically, they argue that though its ancestral home is Britain, English in Nigerian has garnered a domestic flavour and some American spices too (Banjo 1982, Jibril 1986, 1998, Awonusi 2004, Adegbija 2004). Likewise, the conceptual question of whether to adopt ‘English in Nigeria’ or ‘Nigerian English’ has also, much recently, lingered. Essentially, while the former underscores an (ESL) status, the latter is in line with Kachru’s (1982) idea of ‘New Englishes’ (Soneye & Ayoola 2015). As explained in Jowitt (1991), a ‘New English’ would imply an established variety possibly in Phase 4 of Schneider’s model for postcolonial Englishes (Schneider 2007, Section 4.1.4) while the notion of ESL underscores a pedagogical framework – with a strong commitment to promoting native-like structures. But given the requisites for descriptivism and its prospects for NigE, the latter is more feasibly suitable.

So far, these splits appear to be bridging, as efforts are now being re-focused on describing what the speakers do rather than prescribing them a norm. Critical to this turn have been inspirations from the variationist proposal in Brosnahan’s (1958) regional classification of NigE and Awonusi’s (1986) synopsis on the educated speakers’ varieties. Based on these outlooks therefore, subsequent studies have as well contested that since the speakers belong to different ethnics, their accents of English would most likely be heterogeneous (Jibril 1982,
Eka 1985, Jowitt 2007). Studies which correspond with this route have hence relied on factors such as ethnicity and formal education in explaining some of the tendencies to expect from NigE speakers (Walsh 1967, Banjo 1971, Jibril 1979, 1986, Udofot 2005, Jolayemi 2006, Akinjobi 2006, Soneye & Ayoola 2015), thus marking a departure from core prescriptive goals towards a more variationist enterprise.

1.1 The Ebira English Subsystem

There is no uniform accent of English in Nigeria. In fact, the diversity of the different kinds of English is so great that Nigerian English (NigE) is usually divided into several sub-varieties. Based on the observation that the native language of Nigerian speakers of English characteristically influences their accent in English, NigE sub-varieties corresponding to different ethnic groups have been proposed (Gut 2004: 815-6).

Notwithstanding numerous commitments towards defining the ethnic substrata of NigE – mentioned above and discussed broadly in Section 2.4, its concept has been largely reductive and, for most part, misleading. Underlying most previous analyses, despite stirring motivations, is what appears to be a hegemonic focus on the varieties spoken by major ethnics in the country, i.e., those whose patterns are often assumed as universal or representative of NigE accent(s). In an obvious neglect of sociolinguistic complexities that typify the evolution of English in Nigeria, its structures have also been chiefly assessed on the clines of lectal similarity or ‘convergence of educated usages’ (Bamgbose 1995, Banjo 1995, Ugorji 2015: 36). Essentially, this paradigm presupposes a range of structural and non-structural features consistently identifiable for the variety, irrespective of other distinguishing variables within the super-system. This omission is regrettable, particularly in the light of sociolinguistic divergences among the country’s ethno-linguistic and social groups. Sociolinguistically, the so-called decamillionaires, viz., Yoruba, Hausa and Igbo (Brann 2004:9) whose English accents have often been deemed overarching comprise a tiny proportion of the 353 currently vigorous sub-systems of Nigerian languages (Figure 2.1, SIL/Ethnologue 2015). However, in view of their population density and geo-political sectioning, majority of outings on NigE have appeared to revolve around only varieties

---

of Yoruba, Hausa and Igbo English or the Southern/Eastern and Northern accents. Consequently, the proposal such as made in Gut (2004:815-6) has been left unattended.

Similarly, the most consistently assessed of differentiating variables in NigE accents has been education (discussed in Section 2.4.2). In fact, the supposed nexus between education and ethnicity basically informs ethnic categorisations such as Educated Yoruba, Hausa or Igbo English in major studies (Brosnahan 1958, Banjo 1982, Jibril 1982, Jowitt 1991, 2006, Udofot 2003, 2004, Josiah et al. 2012 and Oladipupo 2015). As with ethnicity, while education is conventionally defining in accent variation, especially in the L1 English varieties, the correlation is rarely as strong for L2 speakers (Section 5.7.1). Put differently, not only is the effect of education overly exaggerated in the cataloguing of NigE, it has been recurrent at the expense of accompanying factors including age, gender, job status, number of spoken L1, urbanity, age at exposure, age of exposure, linguistic or co-textual factors, speech styles and so on. In addition to these theoretical gaps is what has been lamented as methodological blandness or the drought of replicable procedures in many studies of NigE system. As highlighted in Chapter 2, and as shown in Table 2.8, such laxities are partly responsible for multiplicity of conflicting accounts – some of which lack strong procedural supports.

Excepting on-going works in NigE phonology, these gaps have been routinely evaded or at best sparsely answered. Thus, coupled with the vocalic appraisal of Ebira English system, this study is also intended at initiating a variationist template – one readily employable in subsequent assessments of NigE accents. The L2 systems of so-called minority ethnic groups in Nigeria have not only been absent in previous accounts, their exclusion does imply they neither exist nor have distinct features of their own. This study foregrounds and remedies this gap, at least for the Ebira English (hereafter EEng) variety.

While no prior assessment of Ebira English inventory has been conducted, the sociolinguistic tutelage under which the variety has so far evolved equally makes it attractive to descriptive efforts. First, unlike in Southern and Northern provinces, the responsibility of European education in Ebiraland during early periods was borne by home-grown Yoruba instructors. Since classroom instruction for school beginners at that time was in the L1, the teachers had little options other than to teach pupils in Yoruba language before transiting unto English in the higher primary classes. As mentioned in Section 2.2, this resulted in some sort of bilingual proficiency – in English and Yoruba, especially among older speakers till date. The accent of English taught to this early crop of students has since remained, and understandably,
passed on to younger generations. Linguistically remarkable, also, is Ebira’s membership of the Nupoid phylum (which ties it to mostly Southern Nigerian features).

Given the usual effect of cultural and geo-political affinity on linguistic behaviour, the Ebira nation has recently been espousing the Hausa North – an affinity more anchored on religious kinship than migration history. Consequently, this study equally measures the extent to which either of Yoruba and Hausa English inflects the EEng system (Section 6.0). In terms of phonological structures, neither of Yoruba or Hausa L1 system converges completely with Ebira vowels. On the whole, the Ebira’s 9 vowel system variegates the 7 and 10 vowel system of Yoruba and Hausa respectively (Elugbe 1983, Pulleyblank 1989, Zubairu & Rashid 2015). In the light of substrate influence, as conjectured in Thomas (2011:148), such uniqueness – as in Ebira’s Advanced Tongue Root (ATR)/Retracted Tongue Root (RTR) systems would be expected to yield a likewise different English inventory – especially one with gradient disparities from its neighbouring varieties.

In terms of methodology, a further departure from the literature in the current study concerns my reliance on acoustic information derived from speech samples as well as the outputs of statistical analysis in determining the vowels’ phonemic status (Section 3.4.2); (see Chapter 5). Following vast traditions of vowel measurements in sociophonetic studies, my results are generally based on the evaluation of F1/F2 normalised values. Section 4.2 discusses the data normalisation procedures employed prior to regression analysis in Section 4.2.1. Given the naturalistic nature of my data, I adopted the mixed effects modelling so as to account for the confounding effects of random and fixed predictors in each run (Johnson 2009). This way, the weight of variance explained by the model and sundry idiosyncrasies associated with individual speakers and word items were disentangled.

Although there are rarely questions as to the fact that NigE is the most divergent on the phonological stratum, the crux of discrepancies widely involves the vocalic behaviour, particularly the monophthongal inventory. Using ethnicity and education (the most prominent variables in the literature), different studies have reported separate representations of vowel clusters, i.e., as to whether they are differentiated or identical. So, while larger inventories have been reported for systems with fewer cases of vowel coalescence, e.g., the Educated Hausa English, a smaller number of vowels have been reckoned for mostly Southern systems (Section 2.5.1 & 2.5.2). As broadly reviewed in Chapter 2, vowels coalescence is by no means peculiar to NigE or the L2 varieties. What is peculiar, however, is its conceptualisation. For instance,
regarding such instances of non-differentiation of vowels in most L2 varieties as mergers or splits becomes problematic – given the linguistic realities characterising them.

Conceptually, mergers presuppose change, i.e., involving the merging of previously unmerged phonemes (Hickey 2004, Ahern 2014). The spread of merger is often due to contact or accommodation between merged speakers and non-merged hearers (Trudgill 1986, Herold 1990). It thus stands to reason the inaptness of merger occurrence in L2 varieties, since it is often an offshoot of dialects in contact rather than different languages. This is however not the case in most African English varieties for which ‘mergers’ have been reported (see Hoffmann 2011). The contact has rather been between languages with different underlying systems i.e., between ATR/RTR & tense/lax distinction (Ladefoged & Maddieson 1996). Though often considered weak or speculative, vowel coalescence in non-native varieties is usually explained on the interference theory (Schmied, p.c.). The interference theory basically entails the speakers’ approximation of their L2 systems based on existing L1 features. In fact, majority of phonological assessment of NigE based on the Contrastive Analysis (CA) or Error Analysis (EA) framework underscore this notion of substrate influences (see Section 2.4). My view therefore, supports that – despite the logic of interference theory, lectal variations among L2 speakers can actually defy the substrate forces, thus evolving differently on the strength of other external variables. In other words, not even instance of non-conforming patterns of speakers is traceable to their L1 (Section 5.4). In such contexts, speakers commonly rely on the feature pool of available systems (L1 and L2) for the realisation of English sounds (Mufwene 2002:47, Section 4.1.1). In this study, though historically identical or differentiated phonemes are neither termed as ‘merged’ nor ‘non-merged’, I rely mostly on conventional methods for determining mergers and splits, as well as vowels’ classification based on Wells lexical sets (Wells 1982).

In the following sections, I present an abridged history of English in Nigeria and in Ebiraland. My main research questions are derived from the broad review of existing literature in NigE (Section 2.7). Given the involved scope of my methodology, I commit the whole of Chapter 3 to explaining the details of the research designs, including major experimental procedures before discussions on the theoretical frameworks, data normalisation and statistical modelling in Chapter 4. The key routines adopted in fitting regression models for the assessment of each vowel group (Section 5.2) is explained at the beginning of Chapter 5. Further analysis in subsequent sections of the chapter begins with KIT & FLEECE (Section 5.2.1), then other historically non-differentiated vowels in the NigE system. The chapter ends with a summary of major effects (linguistic and social) in the patterning of vocalic trajectories. And based on the
findings, Chapter presents the conclusion, the précis of findings, as well as encountered study limitations in Section 6.0, 6.1 & 6. A proposal on subsequent areas of future studies is in Section 6.3.
2 Through Evolution, Diversity and Powers of English in Nigeria

The earliest presence of English on the Nigerian soil dates back to 1553 after the visit of British explorers whose prime quests concerned commerce and diplomacy. The chief interest of the British with the locals at this time, as also in other sub-regions, was mainly slave trade – an enterprise which triggered a desperate need for communication between the slave buyers, sellers and the slaves (Spencer 1971:10, Awonusi 1986:555, Schneider 2007:199). Dealings were generally business-like, and mainly built around trade in items like elephant ivory, artefacts and agricultural products. The introduction of formal Western education, however, did not begin until the early 19th century, following the activities of missionaries. Missionary centres were founded by the Wesleyans, Church Mission Society (CMS), the Baptists and the Sociétés Des Africaines and the Roman Catholic Missions with the intent to preach Christianity as well as strengthen the colonial culture among the natives. Crucial to the success of these schools was the teaching of English, which also became a funding requirement by the colonists.

Steadily, schools and religious outlets sprang across the villages of South-west and East of Nigeria, where the language was taught. The English language grew in prominence as freed slaves from Europe and America who returned to Africa also joined in teaching it at these schools. Other forms of trade established in the aftermath of slavery further entrenched the use of English as the language of business; and rulers in regions of the country where colonial presence was more warmly received were exposed to basic means of communication in the visitor’s language. In fact, evidence of legible record-keeping in English by King Eyamba of Duke Town was discovered and documented in 1797 by Reverend Hope Wadell, a Scottish Presbyterian missionary who worked in the region (Ajayi 1965:89, Ayandele 1966:1-2).

Towards the mid-19th century, the relationship between British and the natives had already begun to form differently with waxing tilt towards religious, political and economic absorption. The British finally gained full occupation of Lagos as its colony in 1861 so as to reinforce its interests. By the end of 19th century, English had won a prime place among Nigerians who had started regarding it as elitist and status-defining (Jowitt 1991, Bokamba 1991, Bamgbose 1995, Gut 2004, Jolayemi 2006, Taiwo 2009).
The spread of English through the provinces was quite unique and gradual. Each region certainly has its story of encounter with the language at its early stage of infusion, depending on how accessible or receptive the conduits of propagation were, namely Christianity, Western education and trade. Little wonder these factors have continued to remain as key variables in the description of NigE (Section 1.0). Missionary presence was more tolerated in Southern and Eastern Nigeria where natives were less suspicious of the foreigners. They however encountered long resistance in the North, which before the colonial advent had embraced Islam. In their quest for converts among the natives, Christian missionaries travelled far through the trenches earlier paved for business activities through mostly Southern villages; aided by interpreters (who were chosen from the returnee slaves). The missionaries themselves also interacted with the villagers, their chiefs and all willing converts in cases of direct contacts. Since sermons were preached in English, its spread was much easier among proselytes and the host communities. Schools were built to endear the new faith to those in these regions and English was the tongue of interaction.

At first, English teaching in schools was not chiefly geared at formal learning, but to provide locals with the means to accessing the Bible, and much after in their local languages. A proof of this goal is Bishop Samuel Ajayi Crowder – a Yoruba slave freed after the abolition of slave trade in 1807. He was educated by the Church Missionary Society (CMS) and ordained as priest. He later rose to become the first African Bishop, who later translated the Bible into the Yoruba language (Taiwo 2009:3).

As a consequence of Christian missions in the South and the East, the regions had much earlier exposure to English than in the North and Middle Belt areas. Awonusi however reports that despite the sameness in contact periods, the schools in South far outnumbered those in the East. He recalls that ‘while the number of schools in the West was rapidly increasing, little progress was reported in the East: out of about twenty secondary schools and teachers training colleges in the South in 1913, only two were in the East’ (1986: 556). This would also later become a significant variable in the spread English and its divergent colourings among its new speakers. Though the early teachers of English across Nigeria were the British who either came as Christian missionaries or colonial administrators; the training institutes in the West did also produce local instructors. These comprised the earliest crop of local teachers who also provided some administrative relief to Britain whose personnel on foreign missions had started dwindling after the First World War. Most importantly, the fresh recruits helped contain the exigencies in expanding schools – a strategy Sir Frederick Lugard
considered very efficient and inexpensive. In fairness to logic, whatever forms of English these ‘teachers’ used or taught in the schools they were deployed were the earliest varieties of NigE. Awonusi (1986:557) recalls:

With the exodus of native English teachers, the teaching of English was left to Nigerians (mainly from the West) whose accent was gradually ossifying in a local direction (Yoruba English). Therefore, the teachers had to rely on textbooks for the correct pronunciation, and so the work of Jones proved invaluable. However, the teachers’ mental conception of the work of Jones, particularly affected tense and lax vowel distinctions influenced the model that was taught, i.e. it resulted in the interpretation of the tense/ lax vowels distinction as that of duration (thanks to the length marks!) rather than quality.

He notes further why ‘it is no surprise, therefore, that the Nigerian English accent has a vowel system that is closer to Yoruba and Efik (the Calabar-based language) than to most other Nigerian languages’.

Meanwhile, efforts to drive through the Eastern parts with missionary education were not completely fruitless. At first, the East was firmly resistant to what they considered an incursion – a poise that restricted religious campaigns to the fringes of South-Eastern planes. But after the pre-emption of opposition in the early 20th century, there was a rapid surge in the growth of schools as the Roman Catholic Mission outsmarted the CMS with a number of 355 schools in the region. Relics of this permeation are still present till today. The Catholic schools had teachers, mostly Irish and Scottish who also served as priests in parallel churches – a factor NigE variationists have often held responsible for ethnic nuances in the accent of Igbo English speakers. Attested cases are the plausible derivations of peculiar systems from Scottish equivalents as in: /ɛr/ and /ær/ adapted as [ɛ] and [ε:] in words like learn, modern, etc; as well as /jʊr/ pronounced in Igbo English as [jua] or [ja] as in your (Awonusi 1986: 558).

On the other hand, English in the North had some peculiar twists. On arrival, they jettisoned advances from Christianity and its Western frills (Adetugbo 1977:95). Until the beginning of the 20th century when CMS established schools in the North Central regions, the entire province was resistant to missionary gestures. It is worth noting that the colonial concept framed for governing the North was peculiar and different from the model for the South, East and West, even after amalgamating the protectorates in 1914. The design, known as indirect rule, was fully implemented on the wheel of the surviving feudal formations – a structure that routinely restricted education to the elites. Since there were hardly missionary schools in the North, the colonists were forced to establish schools which catered mainly for
the big men’s wards. In 1909, the first of these was founded in Kano, and were serviced with finest tutors of native British extraction. This, in Awonusi’s (1986) view, was why students ‘who eventually provided the popular accent in use in the North, [still] use accent that is very distinct from that observed in the South’. Leftovers of this linguistic fall out, he observes, still exist in the unique enunciation of /ʌ/ in some NURSE vowels and the voicing of /ʒ/ in vision in Educated Northern English accent; whereas the South–eastern approximations of same remains as [ɔ] or [ʃ]. However, such exclusive access to education did not continue for too long. Adekunle (1995) quoted in Jolayemi (2006) recounts that:

In 1940, after many years of opposition to the introduction of English into the educational system in Northern Nigeria, the Emirs of six emirates (Daura, Katsina, Hedjia, Gumel, Kano and Kazaure) at a meeting told the Governor, Sir Bernard Bourdillon, that although they recognised the religious importance of Arabic, “knowledge of English is ‘Ci gaba’ that is ‘progress’ (file 31727, KNA). English teaching was then allowed in Elementary three (2006:63).

Besides the roles played by the British schools in the evolution of NigE varieties in the early periods, Jowitt recalls that there were also Western businessmen with Cockney, Yorkshire, Birmingham accents, and many American Peace Corps volunteer in the 1960s who lived among the people (Jowitt 1991). These variables, coupled with other social factors, formed the most primitive bases for ethnic distinctions, which have also defined the grounds of enquiry in some previous studies on NigE phonology. The major NigE accents – namely Yoruba, Hausa and Igbo have distinct primary phonological systems commonly inflective in English. For instance, Hausa consists of five vowels which are phonemically contrastive in length with some approximate forms of English central vowels; Igbo has eight vowels grouped into sets of harmonic pairs; and Yoruba has seven phonemes which also contrast in length. These divergences manifest as carryovers in ethnic accents (Jowitt 1991, Gut 2004: 816).

2.1 Functional Powers of English since Contact

Following the colonial Act that brought education under strict government control in 1882, English was officially sanctioned as the language of formal learning in schools, ‘thereby promoting an assimilationist culture’ (Taiwo 2009:4). Throughout the Southern and Northern protectorates, English progressively garnered influence among the people. As a result, a fair
number of locals who fell in love and sought common pacts with the colonists emerged (Igboanusi 2002:9). These fellows keenly learnt the new language. Just after the Second World War, as agitation for self-rule began, the need to prepare Nigerians who would function as administrators in the evolving state became crucial. More elementary training centres were added to existing ones. Between 1940 and 1954, a number of five higher institutions of learning were created, one of which was the Premier University (now University of Ibadan). In addition, there were funding provisions for the training of experienced Nigerians in the United Kingdom – the home of the colonial power (Odumuh 1987:15). This breed of Nigerians sent abroad on government scholarships were contracted to return home after their studies in order to participate in the running of the country (Jolayemi 2006:4). Their presence and the British officers who continued to be active in civil affairs further waxed the esteem of the RP as a model accent among Nigerians. Some British imperial eyes who lived and worked among the few Nigerian elites up till after independence were drawn from the upper and middle class of the British social hierarchy who routinely spoke the English RP or the Southern British English (SBE) (Gut 2004:815).

Since advent, English has progressively morphed from a colonial language to the country’s lingua franca. Given the absurdity of imposing one of the local tongues on others as the national language, English serves as the ethnically nonaligned means of formal and informal interactions across the country. In sub-Saharan Africa, Nigeria counts as the most strongly anglicised country which has fully embraced English as the language of business, education, legislation, bureaucracy, mass media, and even religion. A considerable number of the country’s total population are estimated to use English extensively (Schneider 2004). So ingrained is English that ‘an average person in the remotest village in Nigeria who has never been to school can boast of a few English words and sentences’ (Udofot 2003:4). In Schneider’s design for post-colonial Englishes, English has successfully ‘nativised strongly, and is still gaining ground at a rapid pace’ among Nigerians. He predicts its transit towards endonormative stabilisation, which in his guess, ‘may just be around the corner’ (p. 210). Beyond its official toga, the language also currently wears a decent emblem of modernity and casual means of communication for the many Nigerians.

English prominence in Nigeria rides on several historical imperatives – which invariably have continued to make debates on language planning particularly problematic. In view of Nigeria’s linguistic heterogeneity, some political scholars have in fact described her as ‘a country of three nations’ (Danladi 2013:4) rather than a sole unit. The amalgam of
ethnically diverse regions of the country by the colonial power in 1914 resulted in the partitioning of three major linguistic groups: Hausa to the North, Yoruba to the South and the Igbo to the East; which have remained the so-called ‘major languages’ among some others. In fact, it would be practically imprecise to specify the number of languages spoken in Nigeria, as studies keep unearthing linguistic dissimilarities between systems hitherto classed as dialects of a language (see Ogunmodimu 2015 on the Àhàn dialect spoken in Omuo-Ekiti, South West Nigeria). As at 2015, Ethnologue lists 527 as total number of languages in the country. Of these, 520 are living and 7 are extinct. Of the living languages, 20 are institutional relative to others, while 77 are developing. 353 are vigorous, 27 are in trouble, and 43 are dying (SIL/Ethnologue 2015).

Understandably, the country’s uneven linguistic diversity (Table 2.1) and concomitant ethnocentric sentiments are strong impediments to an effective language policy. Issues that concern such process are rarely divorced from politics and ethnocentrism. While Yoruba, for instance, has over 40 million speakers within and outside Nigeria, alongside Hausa and Igbo spoken by 37 and 25 million people respectively, Ebira, spoken on the borders of these ethnics in the Middle Belt has just about 1.5 million speakers and Oko, minority language in Kogi State has fewer than 100,000 native speakers. In a space as this, the languages with a very high population of speakers tend to become more powerful and oppressive, especially when they ‘match numerical supremacy with economic and political supremacy’ (Adegbija 1994:70). Thus, the dream of a national language (apart from English) is practically unfeasible or would be fervently opposed by languages threatened by elevating a peer language to official or national status.

Seen as a ‘no man’s language’ within Nigeria and an effectively unifying tool, English has continued to play a number of prominent roles among Nigerians. Going by the 1998 National Policy on Education, the government recommends – in the 6-3-3-4 system – the use of the native language as medium of teaching in the lower primary (which is usually the first three years of the six years to be spent at the primary school). Though both English and a foreign language such as French can be taught as separate subjects during this stage, the core language of interaction would be tongue of the kids’ immediate environment. The situation has however changed since the start of the 1990s. The formal education system has since universally relied wholly on English Language as principal means of instruction from the most elementary levels to advanced stages.
<table>
<thead>
<tr>
<th>State</th>
<th>Dominant langs.</th>
<th>No of langs.</th>
<th>Major langs.</th>
<th>Langs. spoken by less than a million speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abia</td>
<td>Igbo</td>
<td>1 Langs/14 dialects</td>
<td>Igbo</td>
<td>NA</td>
</tr>
<tr>
<td>Akwa Ibom</td>
<td>Annang, Ibibio, Oron</td>
<td>3</td>
<td>NA</td>
<td>Ibibio</td>
</tr>
<tr>
<td>Adamawa</td>
<td>Abon, Awak, Bachama, Fulfulde (Fula), Chamba, Jukun, Hausa, Kuteb</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Taraba</td>
<td>Igbo</td>
<td>119</td>
<td>Hausa</td>
<td>Tiv, Fulfulde</td>
</tr>
<tr>
<td>Bauchi</td>
<td>Fulfulde (Fula)</td>
<td>73</td>
<td>Hausa</td>
<td>Nupe</td>
</tr>
<tr>
<td>Benue</td>
<td>Idoma</td>
<td>10</td>
<td>Hausa</td>
<td>Tiv</td>
</tr>
<tr>
<td>Bornu</td>
<td>Baduna, Fulfulde (Fula), Kanuri, Shuwa-arabic, Effik, Annang, Bokyi, Baduna</td>
<td>4</td>
<td>Hausa</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Yobe</td>
<td>Igbo</td>
<td>39</td>
<td>NA</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Cross River</td>
<td>Bekwara, Ibibio</td>
<td>71</td>
<td>NA</td>
<td>Tiv, Effik, Ibibio</td>
</tr>
<tr>
<td>Delta</td>
<td>Afemai (Yekhee), Urhobo Edo (Bini), Itseki, Ibibio, Hausa, Fulfulde</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Edo (Bendel)</td>
<td>Igbo, Ijo (Izon), Urhobo</td>
<td>31</td>
<td>Yoruba</td>
<td>Yoruba (Pidgin)</td>
</tr>
<tr>
<td>Enugu</td>
<td>Igbo</td>
<td>1</td>
<td>Igbo</td>
<td>NA</td>
</tr>
<tr>
<td>Anambra</td>
<td>Igbo</td>
<td>1</td>
<td>Igbo</td>
<td>NA</td>
</tr>
<tr>
<td>Imo</td>
<td>Igbo (18 dialects)</td>
<td>1</td>
<td>Igbo</td>
<td>NA</td>
</tr>
<tr>
<td>Jigawa</td>
<td>Hausa, Fulfulde</td>
<td>2</td>
<td>Hausa</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Kaduna</td>
<td>Hausa, Kaje</td>
<td>53</td>
<td>Hausa</td>
<td>NA</td>
</tr>
<tr>
<td>Kano</td>
<td>Hausa</td>
<td>4</td>
<td>NA</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Katsina</td>
<td>Hausa, Fulfulde, Arabic, Bare, Badakare</td>
<td>2</td>
<td>Hausa</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Kebbi</td>
<td>Banga</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Sokoto</td>
<td>Fulfulde, Hausa</td>
<td>17</td>
<td>Hausa</td>
<td>Fulfulde</td>
</tr>
<tr>
<td>Kwara</td>
<td>Baruba, Agwara-Kamberi</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Kogi</td>
<td>Ebira, Nupe, Yoruba, Igala</td>
<td>18</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Lagos</td>
<td>Yoruba</td>
<td>4</td>
<td>Yoruba</td>
<td>NA</td>
</tr>
<tr>
<td>Niger</td>
<td>Gwari, Hausa, Nupe, Fulfulde</td>
<td>23</td>
<td>Hausa</td>
<td>Fulfulde, Nupe</td>
</tr>
<tr>
<td>Ogun</td>
<td>Yoruba</td>
<td>2(2)</td>
<td>Yoruba</td>
<td>NA</td>
</tr>
<tr>
<td>Ondo</td>
<td>Yoruba</td>
<td>2(8)</td>
<td>Yoruba</td>
<td>NA</td>
</tr>
<tr>
<td>Osun</td>
<td>Yoruba</td>
<td>1(7)</td>
<td>Yoruba</td>
<td>NA</td>
</tr>
<tr>
<td>Oyo</td>
<td>Yoruba</td>
<td></td>
<td>Yoruba</td>
<td>NA</td>
</tr>
<tr>
<td>Plateau</td>
<td>Hausa, Jarawa, Jukun, Tiv, Ikwere, Izon (Ijo),</td>
<td>99</td>
<td>Hausa</td>
<td>Tiv</td>
</tr>
<tr>
<td>Rivers</td>
<td>Kalambari, Kana</td>
<td>33</td>
<td>NA</td>
<td>Ijo</td>
</tr>
<tr>
<td>Abuja (FCT)</td>
<td>Gwari, Hausa, Nupe</td>
<td>10</td>
<td>Hausa, Igbo, Yoruba</td>
<td>Nupe</td>
</tr>
</tbody>
</table>

Table 2.1: Language distribution in Nigerian states (Adegbija 2004).
As a consequence of linguistic diversity in most Anglophone Africa, ethno-political and national sentiments are by and large core in the way speakers relate with English. The influence of English in a number of former African British colonies remains very strong till date. Hence, the use of imperial languages is deemed the inevitable alternative to conflicts and tensions that might arise from ‘enthroning’ a local language over its counterparts. Thus, the understanding of attitudinal complexities among the natives in such multilingual contexts can inform a genuine basis for language planning or the standardisation of existing ones. And by all means ‘the attitude of individuals towards a particular language may affect language shift or language death in society’ (Baker 1993:5 as cited in Adegbija 1994:65).

Some of the earliest studies on language attitudes among Anglophone Africans are Schmied’s (1985) on Tanzania, Sure’s (1991) on Kenya and Adegbija’s (1994) on Nigeria and sub-Saharan varieties. For instance, one of Schmied’s goals is to uncover how hidden indices of language attitudes among Tanzanians can assist in solving the problems of language policy, use and pedagogy (1985:237), and finds that most natives favour only English as the language of education because it affords all children (of different ethnic roots) equal opportunities for learning. There are some who however feel that English foists a cosmetic identity on African, hence must be embraced with caution.

Attitude towards English in Nigeria are much alike – albeit with some nuances. From the earliest period, the mode of acceptance, use and pervasiveness of English differed from region to region and people to people. As at the period English gained popularity among Southerners through the missionary schools, Christian mission were denied access in the Muslim areas of the North, subsequently curtailing its spread until the establishment of the first European school in Kano in 1909 (Gut 2004: 815). But about this same period, English had become prominent and status-defining in some districts of the South. Bamgbose (1996) reports that the first set of Nigerian District Officers held themselves in fashion of the British visitors, who insisted on English with their kinsmen with whom they share a local language, and also had their words interpreted into appropriate Nigerian languages (p.357).

In fact, parents withdrew their children from schools when the colonial administration tried to complement the language of study with the local languages. The evident badge of education and modernity, for many, is the ability to communicate in English (Awonusi 2004:49, Taiwo 2009:4). This attitude persists till date in most Nigerian homes. Parents who send their kids to non-public schools often insist, despite government policy, on their wards acquiring English right from the crèche level through the elementary classes and making sure
that the teachers in such schools enforce strictly the use of English as the only means of interactions. Since the private pre-primary or the nursery school system has been in vogue across the country, children enrolled in these crèche begin contact with English from cradle. ‘More than 90% of the pre-primary schools in Nigeria do not follow the prescription of the National Policy on Education to the effect that the first languages should be the medium in early education’ (Adegbija 2007: 209). In fact, a good number of Nigerian parents – especially in the cities, mostly of the elite/middle class including parents who do not share the same L1 communicate with their kids in English right from infancy. Aside those linguistically constrained to this trend, others do so with the intent to instruct their kids only in English – a dream inspired by the notion of English being key to social wellness and acceptance. This attitude is clearly couched in Soneye & Ayoola (2015):

The use of the language of the immediate environment is almost a taboo in many of such schools...Children from the middle-lower class and the middle class homes often have access to satellite televisions, computer games, play stations, cartoons, videos, storybooks and other educative materials all produced in English...and used English is naturally used at homes and playground.

There are already predictions that English could become the mother-tongue of many Nigerians by the middle of the 21st century (Udofot 2003a). This, of course should call for concerns, as this reap of ‘L1’ English speakers, on emergence, may end up speaking no language with native-like competence. The interview excerpts in Adegbija (1994) further attest to these realities (Table 2.2): one of a housewife living in Abuja, the Nigerian capital city (Respondent A), and the other, a medical doctor from the Old Bendel State (respondent B):
Table 2.2: Interview excerpt on the attitude and spread of English among Nigerians

Tracing the core of Nigeria’s language policy and difference in the advent as well as the stretch of English among ethnics is fundamental to this study. I propose to provide an empirical support for features already attested in NigE phonological systems, while also accounting for possible variations occasioned by social factors. The missing link in the overall efforts to describe the phonology of NigE is the obvious dearth of procedural designs and variationist outlooks on data. My design for this study thus marks a departure in several ways. First, the focus is on Nigerian ethnic minority whose system has never been studied, or generally fused up with those of the larger groups. Second, the data is structured and gender-stratified according to social categories. Regrettably, given the unique complexity of social stratification among Nigerians, it is difficult to replicate some famous variables often implicated in the assessment of the established varieties in the West African contexts, especially Nigeria.

The choice of sounds or syntax a speaker makes, in most cases fails to truthfully indicate his place on the social cline. This fact regardless, and as far as English is concerned, most description of speakers’ proficiency in Nigeria essentially correlates with their level of education (Banjo 1971, Awonusi 2004, Gut 2004). In addition to level of education, I propose that factors such as age, gender and the kind of job could also have significant effects. My choice of ‘age’ as variable is also linked to major developments in the implementation of Nigeria language policy in schools since the early 1990s. For instance, the older generation of speakers in this study attended primary school during periods of strict enforcement of the government’s language policy – and were not taught in English until the upper primary
classes. During these periods, a child must have spent a minimum of five years with the parents at home before enrolment at schools. Invariably, the onset age of exposure for most speakers in the group was eight or nine. The situation is however different for younger speakers most of whom were already exposed to English at home before kindergarten at the age of 2 or 3, and never had formal interaction in their mother tongue throughout schooling up till the time of data collection. In view of this, the prospect of correlation between onset of exposure and age is anticipated, and it would be interesting to measure these relationships.

2.2 The Ebira People and English in Ebiraland

Similar to the missionary activities, colonial presence in Ebiraland must have stretched from parts of the country which had earlier come under indirect rule. Despite the regal consolidations which began in the 1880s, the first administrative encounter of the British with Ebiraland was not until 1900 – when Sir Frederick Lugard took over the government of Northern Nigeria from the Royal Niger Company and hoisted the British flag in Lokoja, the present capital of Kogi State. Prior to this time, British presence had begun to spread across the Northern borders, and had taken over the administration of key settlements along the Niger-Benue areas: Bida (in the present Niger State), Ilorin (in Kwara State), Ibi, Wase and Donga between 1884 and 1898. But the conquest tide did not stretch down into Ebira nation until 1903 after Goldie received the authorisation of the British Privy Council for the establishment of the Royal Niger Company (RNC) in Lokoja. Many historians believe this as mere smokescreen for further occupation of sovereign Northern territories (Suleiman 1992, Okene 2000).

The major items of trade through the company were kernels, palm oil and cotton from far provinces in the North, and other agricultural products through its operational headquarters in Lokoja. The whole of Ebiraland was later marked out as an economic annex of the RNC in 1890, and military garrisons built in surrounding municipals to forestall uprisings (Abubakar 1980, Okene 2000). With huge military presence in Northern cities, lesser provinces like Ebiraland and other towns within the Middle Belt areas became vulnerable to British’s imperial might. In 1890, Governor Lord Lugard, a British explorer and colonial administrator took over from the RNC and stationed his base in Lokoja. His style of conquest differed somewhat from his predecessors’. To conquer the land and fully have it under his rule, he admixed diplomacy with force. Upon arrival, eight of his men led by Captain Beddoes were
sent to explain the terms of relationship to the Ebira nation – a proposal flatly rejected by the chiefs (Willis 1972, Okene 2000). In 1902, a small military escort led by Mr. Malcom and Lt. F.F. Byng-Hall were sent to explore and subject the area under colonial rule. As it was in other regions, the people resisted until they were subdued by the superior fire-power of West African Frontier Force (WAFF) pitched amidst the locals to ‘tame’ them. By 1904, Ebira finally agreed to co-operate with the English under Mr. Morgan who at the time was the Resident of Kabba and all Ebira provinces.

The British conquest of most African territories rode on the back of evangelism. The earliest means of contact with English in most African cities was through the missionary activities – later hijacked by the colonial powers. Natives who had embraced the new faith were promptly taught the basics of English and made local evangelists to help make proselytes of their kinsmen and equally teach them the new tongue. A similar scenario marked the arrival of English in Ebiraland at the early stage. The majority of the early schools missionary centres mostly run by teachers who were either mission clergies or attachés of the religious movements. They preached and taught in English, both in churches and schools. Local missionaries who came from the South also relied on English and Yoruba (being their first language) for their assignments. As fact, the first pupils of the CMS elementary school in 1918 were the church’s Catechist and Festus Alusoka – who both were of Yoruba extraction. Though some Arabic Jihadists – who had earlier invaded from the North – settled among the natives, their linguistic influence was relatively slight. Proficiency in Arabic was available to all, but largely a product of special learning from Imams and Islamic scholars at seminaries. On the whole, Christianity has played a central role in the introduction and the use of English in Nigeria from the earliest time till date. Most Pentecostal churches across Nigeria, large or small have continued to conduct their services in English till date, with some make-shift arrangements in native languages for those who may not be able to follow. 70% of the most attended elementary ‘good schools’ in Ebiraland are private-owned and thus enforce the enviable ‘strictly English speaking’ tradition, making the once visitor’s language universally ubiquitous. The Hollywood/Nollywood revolution, and the general media are equally significant agents of linguistic in-breeding, as movies, clips and broadcasts in English and Yoruba are amply available as routine staples to all.
2.3 The Benue Congo Phylum

Ebira, Nupoid, is of the Kwa Benue-Congo group spoken by over 1.5 million people as a minority language in Nigeria (Greenberg 1963, Hansford et al 1976, Adive 1989). The Ebiraland comprises the hub of Ebira Tao which makes up the central districts of Kogi State in the North Central of Nigeria (Figure 2.1). Its geography stretches about 23 kilometres west of Niger at Ajaokuta and 32 kilometres southwest of the Niger-Benue confluence, with a landmass of 3,426 km square hedged by hilly undulations within the Middle Belt region. The core of the Ebira ethnic group currently live in six major local administrative units: Adavi, Okene, Okehi, Ajaokuta, Ogori-magongo and Lokoja, with ethnic annexes in other states across the country – in Ondo, Oyo, Osun, Nassarawa, Edo, Benue State and the Federal Capital Territory. Their ethnic history dates back dimly to 1680 when they migrated from the Kwararafa confederation (the present Taraba State) across the deserts of the North to settle near the River Niger, forming one of the ethnic groups in the Middle Belt. Apart from Ebira, other main Nigerian languages in the Kwa group are Yoruba and Igbo spoken in the Southwest and East of Nigeria (Adive 1989:6); some of which are its geolinguistic neighbours: Yoruba, Igala, Edo and Nupe to South, East and North respectively.

Among languages surrounding Ebira, Yoruba appears to have exerted the most influence on Ebira through a number of events (Adive 1989:2). He recalls that since Western education and missionary activities spread from Western Nigeria to Ebiraland during early contact, the logic is plausible. This is true considering the origin of formal education in Ebiraland. The teachers sent to CMS and Roman Catholic Missionary schools established in these areas at the start of the 19th century were mostly Yoruba instructors from the South, and the medium of interaction in the lower primary classes was Yoruba – a situation that occasioned many Ebira natives being bilinguals in the two local languages till today.
Figure 2.1: An index map for some Nigerian languages. Ebira (302) is mainly spoken around Kogi/Lokoja area. Nupoid cluster can be found within the areas marked in spring green (source: SIL/Ethnologue 2015)
Another factor is the mass migration of the Ebira people to the South in search of arable lands for farming, owing to the aridity of their home topography for major agricultural activities. Most Ebira natives who were living in Yoruba regions unsurprisingly became proficient in Yoruba, and some even lost their native tongue to it. Yoruba has about 40 million primary and secondary speakers across Nigeria and millions of speakers in some areas of Benin and the Americas making it the most widely prominent African tongue outside the continent. It yields a strong areal\(^3\) influence over a number of languages now classified as belonging to the Yoruboid tree such as: Igala, Ogugu, Ede, Itsekiri, Yoruba and Olukumi – a lone variety in the Edoid\(^4\) family.

### 2.4 Ebira and Yoruba Vowel Systems

Aside two key accounts of Ebira structures – particularly the phonotactic tendencies, harmony patterns and syntax (Adive 1989: 11-43, Orle 2003), very little have so far emerged on the phonology of Ebira. The two influential accounts in Adive (1989) and Orle (2003) are the most persuasive outings in the literature up till the time of this study, thus my reliance on them for the review of the L1 structures. Yoruba, however, has enjoyed much linguistic attention and attracted layers of documentation, particularly in works on African phonology. I find these languages (Ebira and Yoruba) mainly crucial to the understanding of phonemic behaviours in EEng, and the understanding of my research questions. Thus, the summary of the Ebira ATR\(^5\) features in this section is based on Adive (1989), Orle (2003), and of Yoruba on Orle (2003), Akinlabi & Liberman (2000) and Pulleyblank & Archangeli (1994).

The standard Ebira has 9 vowels, 8 of which are in ATR/RTR pairs, leaving the low central [a] as neutral; while the standard Yoruba has 7 seven oral vowels with non ATR contrast in high front and back. For instance, Ebira has /–i/ and /+ i/ & /–u/ and /+u/ but the Standard Yoruba only has the advanced variant of the two, i.e, /+ i/ & /+u/. Noting this

---

\(^3\) The areal component of languages underscores strong linguistic similarities between two or more languages across levels of representations, ranging from a common genetic history or descent, to translingual borrowings. It could also a consequence of feature retention among the population adopting a new language, or simply owing to chance (Campbell 2006).

\(^4\) Edoid languages consist of thirty six languages spoken mainly in some Southern parts of Nigeria (Elugbe 1989a, 1989b).

\(^5\) Advanced Tongue Root (ATR) and Retracted Tongue Root (RTR) are characterised by the either expanding or contracting the pharyngeal cavity. Thus, vowels produced with raised tongue root have lower F1 and higher F2 (higher on the y-axis, and fronter on x-axis), while retracted ones are relatively higher in F1 and lower in F2. ATR patterns, i.e., in F1 & F2 however differ cross-linguistically.
distinction is decidedly crucial to this study, considering the sisterhood of both languages and their contact situations which are perhaps consequential in the overall constitution of EEng (see 1.1 above on contact events between the ethnics). Apart from tonal spread\(^6\) in Ebira, both languages have three phonemically distinctive and lexically contrastive tones. Their vowels and nasals are mostly syllabic, and apart from some instances of ill-formedness\(^7\) in high vowels, both systems freely participate in the ATR/RTR mechanisms (Table 2.3). Long vowels are marked as germinate in some Nigerian languages, i.e., two or three of same vowel can occur within a syllable, e.g.: \(\text{àåhè} – \text{play or drama}, \text{òòhè} – \text{free of charge}, \text{ùùhwè} – \text{hen}, \text{èèva} – \text{two or three as in grammatically complex word àáávé – if he/she}^8 \text{ is coming, etc.} \) (Adive 1989: 12&18). Such occurrences, however, do not correlate with length as in English long or tense vowels. Within the identical clusters, it is possible to assign different tones, ‘and because tone is a distinctive feature of the syllable,..., a germinate vowel sequence spreads over two syllables’ (Adive 1989: 17). Unlike in English diphthongs, for instance, different phonemes cannot occur within a syllable, but Adive considers the second item in the germinate sequences as schwa (\(\text{ə}\)) – a feature most improbably attested for Nupoid forms.

\[
\begin{array}{ccc}
\text{+i ATR} & & +u \text{ ATR} \\
+\text{high} & -\text{i RTR} & -u \text{ RTR} \\
-\text{high} & +\text{e ATR} & +o \text{ ATR} \\
-\text{e RTR} & -\text{o RTR} \\
\end{array}
\]

Table 2.3: Ebira vowel harmony sets

Drawing on Carnochan (1970:224)\(^9\), Ebira manifests what may be described as ‘structural schwasyllable prosody’. By auditory evaluations, the second phonemes of the

---

\(^6\) I use the term ‘tonal spread’ to denote instances of kinetic tones such as the high-falling tone and low-rising tone in the final syllables of \(\text{ùò} - \text{cowrie shell for divination}, \text{òdà} - \text{law}, \text{etc.}\)

\(^7\) I use this term to denote the opposite of well-formedness (Pulleyblank & Archangeli 1994:8), which in the Modular Phonological framework encodes the universality or applicableness of rules and constraints alike in the structure of different languages.

\(^8\) Ebira has no pronoun marker for gender.

\(^9\) Carnochan in his study of Igbo vowel system reports that ‘the vowel sound in the second syllable of each example is the same as in the final syllable; together they constitute one alternance.... The V-\(\text{ə}\) phonological notation indicates the interdependence of the syllable: \(\text{ə}\). This V-\(\text{ə}\) phonological notation indicates the interdependence of the syllable and correlates with hearing the same vowel sound in both syllable and
germinate pairs share close phonological similarities, thus ruling out the possibility of schwa realisation in those contexts. And as a native speaker, this would likewise be my impression, especially when the germinate sequence is no consequence of consonant deletion. Also, major studies on NigE using the Error Analysis (EA) framework would fault Carnochan and Adive on the assumption of schwa realisation in these languages (cf. Akinjobi 2009:93, Josiah et al. 2012). Absence of weak forms such as the schwa is often marked for the Southern variety of NigE, plausibly due to the perceived loss of a similar feature in the native languages (Gut 2004, Awonusi 2004).

2.5 NigE Varieties: Ethno-linguistic Nexus

It usually takes only a few utterances by a speaker for a listener to make educated guesses about his or her regional, and to some extent also, social origins – our accent betray where we are from (Schneider 2004:72).

In the light of evolutionary complexities in NigE since contact, identifying the social determinants responsible for varietal distinctions could be a knotty task. Unlike for mostly inner circle varieties- especially in the West, conjecturing patterns of correlation between social factors and lectal nuances among NigE speakers is usually problematic. Given the ethno-linguistic diversity and uneven access to formal education, the often-named of such variables include ethnicity and education (Brosnahan 1958, Banjo 1971, Bamgbose 1982, Udofot 2003, Gut 2004, Olaniyi 2014). NigE phonologists maintain that insofar there are dissimilar structures in the host languages and disparate levels of educational trainings, different patterns would be expected in the L2.

Understandably, the link between ethnicity and languages in Nigeria is likewise complex, as scholars vary on the sociological and linguistic entailment of the term. While Hansford et al (1976) put the number of linguistic groups in Nigeria around 400. The records of Ethnologue in 2015 reviews it to 527 following continued discovery of more minority languages and hitherto dialects recently assessed as linguistically independent (Section 2.1.1). From a sociological standpoint, Ugbana (1977) profiles the ethnic groups at over 250. The question thus is deciding the more ideal of these parameters in terms of categorisation, i.e. ethnicity or linguistic groups?

Opting for a taxonomy based on linguistic variables rather than ethnic groupings seems more promising, particularly if the focus is not on factors external to language structures and the correlates with hearing the same vowel sound in both syllable (Carnochan 1970:224)” as cited in (Adive 1989:18).
intent is to describe their distinct features. But the literature on NigE has so far dealt differently. Varietal classification is often traced to a very few main ethnic sources – especially the *decamillionaires*, namely Yoruba, Hausa and Igbo English (Brann 2004: 9). A lot has been published on these sub-forms, thereby entrenching them in major discussions on NigE. There are a few reasons this is so: first, the three groups have represented Nigeria’s geopolitical strata since the amalgamation of 1914\(^{10}\); and second, they have continued to remain notably dominant over other ethnics in the regions (Table 2.4).

<table>
<thead>
<tr>
<th>Category</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regions with single dominant language</td>
<td>South West (Yorubaland), North East (Kanuri-speaking), North (Hausa), South Central (Igbo), South East (Ibibio)</td>
</tr>
<tr>
<td>Regions of high diversity</td>
<td>Northern Cross River, South West of Niger Benue Confluence, Jos Plateau, Muri Mountains</td>
</tr>
</tbody>
</table>

Table 2.4: Ethnodemographic patterns in Nigeria (Blench & Dendo 2003:3)

While Yoruba speakers lead in the whole of South West, Hausa is ubiquitous throughout the North and Igbo in South Central and East. Forming a parallel with Kachru’s concentric circles, Olaniyi (2014:236) draws a pyramid of NigE to mark the density gradience of each group and their classifications. Based on the listing of Nigerian languages in Adegbija (2004), Olaniyi proposes a possibility of three different varieties of NigE. On the top of the schema are the ‘big three’, due to their huge influence in Nigeria’s linguistic space. Citing Udofot (2004), he notes that these accents ‘have in the literature of Nigerian English earned the nomenclatures of Hausa English, Igbo English and Yoruba English as sub-varieties of Ninglish’. In the lower circle are major minor languages, i.e., languages with more speakers relative to much lesser ones: ‘Fufulde, Annang, Ibibio, Edo, Tiv, Kanuri, Ebira, Nupe, Bassakome, Idoma, Itsekiri, Igala, Urhobo, Ekiti, Igbomina, Ijebu–Ijesa, Egba, Gwari, Ilaje, Ondo, Agatu, Idoma, Kananci, Efik, among others. The lowest rung consists of over 150 languages whose speaker population are fewer than those of the lower circle. Ihiala, Nnewi, Ogbaru, Nsuka, Idemili, Oron, Abon, Awak,

---

\(^{10}\) In 1914, the Southern and Northern Protectorates were brought together as one colony under Sir Lord Lugard, forming a sole administrative unit for the regions.
It is important to note, however, that some of the languages listed in the lower and lowest circles may also be considered as ‘radically distinct’ dialects of the ‘big three’, or too structurally dependent for a unique status (Figure 2.2). To further justify the groupings, Olaniyi explains that while those of the lower circle have received some form of linguistic attention, there is still dearth of codification efforts on the lowest systems. The prospect of such NigE taxonomy as in Olaniyi (2014) however faces a serious challenge – bearing in mind the heterogeneity of languages listed together as members of a similar variety. And despite theoretical elegance of Kachru’s model he adapted, the grouping is neither based on genetic resemblance nor geolinguistic proximity; but chiefly on the size of speakers’ population.

Figure 2.2: Olaniyi’s pyramid for NigE varieties (Olaniyi 2014:237)

2.5.1 Level of Education

The earliest correlates of formal learning for NigE is in Brosnahan (1958), who in his article, *English in Southern Nigeria*, reckons speakers’ level of exposure to formal education as mainly predictive. Gauged against the school system of the time, he bands up the varieties into four major groups, which roughly correspond to the existing tiers of education. But his design also includes speakers who never had any formal training with the language, but seldom use it for communicating with folks. In group one are those who picked it outside formal system, among
whom are competent Pidgin speakers. The second group comprises speakers who did not proceed beyond the primary school level. Those in the third and fourth groups already have secondary and university degrees respectively.

Subsequent studies have however dismissed this model as faulty, querying the place of Pidgin among the outlined varieties of English (Mafeni 1971, Banjo 1971, Platt & Min Lian 1983, Elugbe & Omamor 1991). Unfortunately, these groupings have been essentially impressionist, hence difficult to assess against successive studies based on similar variables. Going forward, the pick of education as chiefly indexical of variation has also been disputed, especially in the light of other confounding factors external to formal school trainings (Bamgbose 1982; Udofot 2003; Gut 2004). Banjo (1971) recommends a similar ranking – more largely on linguistic variables than speakers’ level of education. His design contains a grouping of NigE accents into a subset of four. The first variety, mostly spoken by menial workers ‘whose knowledge of the language is very imperfect’, and for whom ‘English is in effect a foreign language’, while the speakers in his group II demonstrate remarkable intelligibility to their fellow Nigerians and to international ears as well. The third are heard among 10% of Nigerians who share the RP flavour at the deep structures but sound ‘Nigerian’ on the surface features. The fourth group consists of few Nigerians who either acquired English as their L1 or have lived long enough in English environments. He however notes the proneness of group four speakers to social stigma from the local audience, especially those who scorn it as cosmetic or markedly different from common flavours.

In Banjo’s (1993) evaluation of the category of speakers in tier one, he agrees with Broshanan’s categorisation mainly on the logic of differentials between English vowels and that of most Nigerian L1 (e.g., many Nigerian languages operate a seven-vowel system /i,e, ɛ, a, o, u/ while English has about 23 vowels). This, he notes, often results in the coalescence of: /liv/ with /li:v/ (as in: leave and live); /kɔt/ with /kɔ:t/, /kət/ and /kə:t/ (as in: court, cot, cut and curt); /wɔk/ with /wɔ:k/ and /wɔ:k/ (as in: walk and work); /kat/ with /kæt/ and /kaːt/ (as in: walk and work); and schwa as [a]. With regard to consonants, the absence of contrast between /z/ and /s/ is reportedly common for Yoruba English speakers, often resulting in homophonous overlap between sue and zoo, for example. Similarly, he suspects that speakers of some Yoruba dialects are inclined to realise /l/ as /v/ and /ʃ/ as /tʃ/.

For Igbo English speakers in this group, Banjo indict the native vowel system for the imposition of ‘vowel harmony in a word like follow, pronouncing it as /folo/, while Edo Variety 1 speakers enunciate the nasalisation patterns of their mother tongue in words like when,
pronounced as /wɛn/” (p. 265-6). *Variety II* speakers are those who demonstrate some awareness of basic distinctive behaviours between the phonemic systems of their mother tongue and the L2. Though speakers tend to collapse more of their mother-tongue features with English, the inflective weight is lesser than for members of *Variety II*. Banjo infers that ‘since this variety is learned at local primary schools, it qualifies as a range of diatopics’; such that, as in *Variety I*, it would be meaningful to talk, for example, of Hausa *Variety II*, Igbo *Variety II*, Edo *Variety II*, Efik *Variety II*, Yoruba *Variety II*, etc’. While speakers in this category have no difficulty in the pronunciation of peripheral phonemes around the vowel envelope, they have difficulty with the realisation of central vowels. The production of plosives and fricatives is similar to RP, but cluster simplification is perceptibly typical. *Variety III* speakers fill the upper space of the continuum as what they speak assumes more phonemic distinction and shows clear departure from those of the former groups. The common phonological structure of this group (apart from subtle phonetic detail which may be missing) apes the RP. Thus, for example, a long monophthong /oː/ equals to RP /ɘʊ/ as in GOAT monophthongisation, while the equivalent of /æ/ is more central in articulation. Similarly, the long vowel /eː/ is same as in RP and some other native varieties of English, i.e., the Scottish and American standards. He observes no profound differences in the consonantal articulation for speakers in this group. Those of *Variety IV* are speakers who were either brought up by the native speakers of English or raised in the inner circle countries where English is mother tongue.

The uppermost tier in Banjo’s classification (*Variety IV*) has come under suspicion in terms of how faithfully representative of NigE it is. Bamgbose (1982) suggests its removal, but Banjo insists on its inclusion in the categories, albeit as the refined form of *Variety III* (banjo 1993:272). Another follow-up on lectal groupings is Udofot (2003) who has also demanded the exclusion of the topmost variety in Banjo’s model, and instead proposed a 3-tier model which, in agreement with Brosnahan, correlates English proficiency with the speaker’s education. Gut (2004) abridges it:

The Non-Standard variety has distinct segmental and non-segmental features such as lack of fluency, an abundance of pauses, a restricted intonation system, and a distinct speech rhythm and accent placement. It is spoken by primary and secondary school leavers, holders of NCE (Nigerian Certificate of Education), OND (Ordinary National Diploma) and some University graduates. It is the variety used by primary school leavers. The Standard variety has a distinct phoneme inventory and characteristic prosodic features in terms of speech rhythm, intonation and accent. It is spoken by

---

11 Diatopic in linguistics refers to geographic indices in language variation.
university graduates and lecturers and other professional as well as final-year undergraduates of English, secondary school leavers and holders of Higher National Diplomas. The Sophisticated variety is spoken by university lecturers in English and Linguistics, by graduates of English, the Humanities and Mass Communication, speakers who had some additional training in English phonology and those who spent some time in English native-speaking areas. It is different from British English in some phonemes and some aspects of speech rhythm, intonation and accentuation (p.817).

2.5.2 The Lects and Quest for Standardisation

Evident shades of structural difference between varieties are often referred to as ‘lects’, basically on the most to least standard forms (Figure 2.3). The terms: *acrolect*, *mesolect* and *basilect* which correspond to highest, middle and lowest variants on the cline was first used by Stewart (1965:15) in his article *Urban Negro Speech* in typifying the continuum of registers within speaker’s linguistic ecology. On the topmost of gradation is the most formal and prestigious form referred to as *acrolect*, followed by *mesolect* and *basilect* occupying the intermediate and lowest rungs.

![Figure 2.4: An illustrative lects ladder for *I am eating* in Jamaican speech continuum (Sebba 1997:211)](image)

Though the theory is more fitting on creoles, pidgins and mixed languages (especially for Jamaican Creole types), it is also suitable for modelling variations in L2 and generally in contact languages (see Bickerton 1975a, Sebba 1977). Further stressing its theoretical validity, (Velupillai 2015) suggests that the sort of nuances *lects* can capture ‘are not
particular to pidgin and creole languages, but can be thought of as a continuum of ‘lects’ that are more or less similar to a given standard language...a non-discrete, continuous range of variation where there are no sharp borders’ (p.212). In the analysis of post-creole continuum, it is possible for a lexifier to retain its influence, even long after the creolisation process is completed – a result of which would be the acrolectal forms. Also, the basilectal variants usually emerge as an offshoot of decreolisation, which also manifests as clear modification of its lexifier. The basilect, on the other hand, usually retains all its original features, and is usually regarded as colloqual and least prestigious among the speakers’ community. In such contexts, lectal classes are by and large continuous with no clear-cut separation between them, thus making access to more than one lect possible, depending on the sociolinguistic requisites.

As part of wide-ranging accounts of social characterisation in NigE systems, studies have either referenced or adapted the lects model for clines of structural grouping and variation indices for the variety (see Banjo 1993, Bamiro 1994, 2006). Such taxonomy has prominence in phonological studies, i.e., where speakers get lects membership depending on their accents or perceived oral competence. A reasonably broad example is Ugorji’s (2008, 2010), who in line with Stewart (1965), categorises the NigE accents into: basilect, mesolect and acrolect. Contrary to Banjo (1993), Udofot (2003), Ugorji believes that these layers interlock in a way that makes clear-cut demarcation of the groups difficult. The basilect nears the accent of those who have a minimum of primary education and matches somewhat with what Banjo (1993) tags as Variety I&II speakers. Mesolect is spoken by the Nigerian parallel of High School students or those who have been trained in such schools, while acrolect marks the highly educated and those who may be particularly trained to speak the language.

According to these categories, speakers of the basilectal variety have seven monophthongs (pure vowels) FLEECE /ɪ/ in live, leave, sit, seat – with KIT as seldom variant. DRESS /ɛ/ vowel as in bed, wet, weight, dead, etc, with possible occurrence of /ɛ/. /ɜ/ is used in the monophthongised variant of SQUARE and NEAR in tear, fear, fair, dare, wear, etc. Low vowels as in lettER, TRAP & BATH are all merged and realised as /a/ in words like ado, about, comma, hat, staff, hart, etc. The low back THOUGHT and LOT are not distinctive for this group, and /ɒ/ is realised perceptibly open for speakers in Northern regions. The absence of diphthongal glide in GOAT, for instance spreads the short /o/ to words like go, bone, dope, dose, etc., with the possibility of extreme lowering of the sound to /ɒ/ in the North. He implies differentiation of FOOT from GOOSE as in do, food, good etc. The complex vowels for these speakers are PRICE /aɪ/ as in bright, kite, bite, tight, etc.; CHOICE
Based on Ugorji’s (2010) catalogue (Table 2.5), the mesolectal variety, for the most part, has the same system as in the basilect, except that some speakers in this group occasionally achieve distinction between KIT and FLEECE. For acrolect, the high front vowels are likewise contrasted, while FACE /eɪ/ gets reduced to monophthongal DRESS /ɛ/. Though what /ɛ/ represents for this variety in Ugorji’s design seems unapt, especially for NigE, the symbol represents DRESS as well as the monophthongal alternants of FACE (as in day, break, take, bake, etc.) and DRESS (as in weight, wet, dead, etc.). Speakers in this group have the unchecked [ɜ] for NURSE, e.g., bird, serve, birth, etc., with /ɛ/ as variant phonemes for TRAP and BATH. Also, their realisation of /ɔ/ and /o/ are much identical to mesolectal equivalents, but could also be heard as /ʌ/, /ʊ/ and /ʊə/ in luck, love, rope respectively – by a highly educated few. The cluster of high back GOOSE and FOOT is very alike the mesolectal variants. Based on these impressions therefore, Ugorji reckons a monophthongal system of twelve vowels for both acrolect and mesolect, including the monophthongisation of FACE and GOAT for mesolectal usages only.

<table>
<thead>
<tr>
<th></th>
<th>basilect</th>
<th>mesolect</th>
<th>acrolect</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple vowels</td>
<td>7</td>
<td>8</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>complex vowels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diphthongs</td>
<td>5</td>
<td>5 – 6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>triphthongs</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>consonants</td>
<td>21</td>
<td>21</td>
<td>24</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2.5: Ugorji’s (2010) summary of Nigerian Received Pronunciation (adapted from Jowitt 2015)

Interestingly, Ugorji’s goal transcends a structural description. He seeks also to prescribe an inventory for a similitude of ‘Nigerian RP’, thus specifying phonemes most suitable for the list and those fit for exclusion from the system. The goal, he notes, is anchored in Schneider’s Model – which describes the NigE variety as actively in transition to the endonormative phase (Schneider 2007). The model acknowledges that despite English’s functional grid in Nigeria, its structure is still largely unstable and heterogeneous12; and

12 According to Schneider’s model, NigE as an Outer Circle variety (Kachru 1985) is in Phase 3. This stage is transitory and structurally predictive. The once colonial language gradually peels off its colourings and adapts itself to the local forms due to conscious tilt approximate its internal
however predicts a gradual progression towards standardisation and pan-ethnic consistency for the variety.

Ugorji’s broadest critique yet – is Jowitt’s (2015) which extensively appraises the highs and lows of the proposal. While the effort draws impressiveness for its application of Optimality Theory (OT), which permits the use tableau and other tools in choosing the ‘fittest’ candidates; he queries Ugorji’s lectal delineation into two major layers, which Jowitt reads as essentially suggesting ‘that we should be thinking, in place of three varieties, of just two varieties – of two polarities – which could be termed “high” and “low”’ (Jowitt 2015:8). He further points out the blurry demarcation of sub-regional varieties in the North, thereby ignoring the role of ethnic diffusion on accents heard in those regions; a taxonomy also implied in Gut (2004: 815-6), Jibril (1986) and Jowitt (1991) in the cataloguing of NigE accents. Most importantly, Jowitt (2015) concludes that the procedures on which such findings and suggestions were based are largely ambiguous and lacking in empirical honesty.

In further defence of his critique, Jowitt presents an auditory account of about ‘an hour lecture given by a Management Consultant whose first language is Yoruba’. In his assessment, he notes that the speaker in question reasonably achieved the schwa /ə/ within some multi-syllabic word-finals; as well as made noticeable distinction between the short and long [a] in items such as transcript, crash, staff and pass. For NURSE, /ɜ/ and /ɑ/ were used instead variants. Another was a speech at a book launch during which he monitored: ‘(a) a Bishop (mother-tongue, Yoruba); (b) an Emir (mother-tongue, Hausa or Fulfulde); (c) a Barrister (‘minorities’, Taraba State); (d) a retired Permanent Secretary (‘minorities’, Adamawa State), and (e) an Educationist (mother-tongue, Igbo)’ (p. 42). For these speakers, GOAT was monophthongal, while schwa-like approximation in unstressed contexts was distinguishable. The stressed variant of [a] was however marked for the Igbo speaker whose accent also indicated the coalescence of /u/ and /el/.

Strikingly, Jowitt ends his critique of Ugorji by proposing that ‘more wide-ranging and more professional monitoring (of his kind of data), especially in the form of a corpus, would surely serve to demonstrate the reality of ‘Nigerian RP’ and its characteristics’ (p.42). First, the prospect of an ‘RP’ for a variety as NigE is not only unfeasible, its concept would generally seem odd, especially in the light structural diversity occasioned by mostly external constitutions among its new host speakers to. Endonormative stabilisation which ushers in the presence of the local norms begins in phase 4. The last phase is when English has fully detached from its source and assumed a new identity – defined by intra-national polarisations and lectal differentiations (Schneider 2003, 2007, Gut 2012).
factors. Also, the extent to which both findings (Ugorji’s and Jowitt’s) explain the speakers phonetic behaviours awaits further assessment, since they appear wholly anchored on auditory, or impressionist judgements.

### 2.6 Overview of NigE Vocalic Inventory: Monophthongs

Two major studies seemed to have opened the grounds for systematic assessment of NigE phonology. The first is Schafer (1967) on the general aspects of speech patterns, followed by Afolayan (1968) whose focus is even more diverse as it also reports instances of lexicogrammatical interference for the speakers. Using existing parameters, they limit their studies to learners of English living in the Southwest. Schafer’s design consists of 13 Yoruba native pupils in their final year. He used about 2000 items of wordlist, some of which were prompted with pictures, colour charts, tables and so on. His findings confirm strong assimilation of the English vowels to five classes with the mergers of high front [FLEECE & KIT] and back [GOOSE & FOOT] vowels. Central vowels are also merged with the nearest classes on each peripheral dimension; and length distinction is absent for checked vowels i.e., merger of [TRAP, BATH letter & NURSE] at the low central. The low back [LOT & THOUGHT] and the mid central vowel [STRUT & NURSE (as in hurt, church, etc.)] are coded as allophones. He reports the causal influence of the L1 in their approximation of [FACE & DRESS] and [GOAT & THOUGHT] thus making it the first account to have hinted on the influence of ATR substrates in NigE accent (Schafer 1967:13-14). On the occurrence of non-differentiation, Afolayan notes the structural divergence between speakers’ L1 and L2 and reckons the absence of central vowels and tense/lax contrasts in Yoruba as mainly reflexive on their accent of English. Part of his major examples include the instances of nasal inflection on [KIT & FOOT] as in tin /tin/ and soon /sùn/ – which according to him, signal indications of host feature transfer (Afolayan 1968:631).

About the same period, Nuttall (1961) also investigates realisation patterns in Hausa English. He reports, as part of his findings, the pervasive switch of short for long vowels and vice-versa in open and closed syllables; and the absence of quality distinction for all central and low vowels. These studies, however, are basically prescriptive, and confined within the error analytical framework – in that such observations were classed as ‘problems’ rather than variation. Consequently, the relevant social factors that might be responsible for these patterns
are not included in the designs. The following summary of Nuttall in a follow-up study by Tiffen (1974) foregrounds this omission:

...Hausa English speakers have problem not only with length but also with quality. Hausa short /a/ has a large number of variants. These are confused with the phonemically distinct English vowels /æ/, /ɛ/, /ʃ/, /ɜː/ and /ɪ/. The most frequently confused are /æ/ and /ɛ/, /æ/ and /ʃ/, /ʃ/ and /ɜː/. The central vowel /ɜː/ is a major problem while /ɔ/ is often given a spelling pronunciation (p.36).

As argued in Section 2.1, earlier studies on NigE phonology have been largely prescriptive, defining deviations from RP as erroneous and suggesting ‘correct’ forms of enunciation to speakers. The most systematic record of this approach in the literature is Tiffen’s (1974; Table 2.6). Like many others, his grasp is fairly universal, covering the segmental and prosodic aspects all at once. The study, however, is seminal and holds many prospects for the current study, especially in terms of the sociolinguistic framework employed. Tiffen recorded 24 speakers for intelligibility tests – 12 Hausa and 12 Yoruba natives, all of whom were in their first year at a federal university. In line with variationist procedures, his data was pooled from wordlist, reading passages of prose and connected speech (interviews). All participants were selected on the proof of being freshmen whose accent had not been influenced by the university’s expatriates, and that they were indigenes of particular ethnics who had no preformal exposure to English at home.

<table>
<thead>
<tr>
<th>Vowel contrast</th>
<th>Yoruba Errors</th>
<th>Hausa Errors</th>
<th>Total Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ʌ, ɜː/</td>
<td>86</td>
<td>50</td>
<td>136</td>
</tr>
<tr>
<td>/ɔː, ɔ/</td>
<td>87</td>
<td>44</td>
<td>131</td>
</tr>
<tr>
<td>/o, œ/</td>
<td>61</td>
<td>53</td>
<td>114</td>
</tr>
<tr>
<td>/ɔː, ɔ/</td>
<td>91</td>
<td>21</td>
<td>112</td>
</tr>
<tr>
<td>/o, u/</td>
<td>58</td>
<td>49</td>
<td>107</td>
</tr>
<tr>
<td>/e, æ/</td>
<td>48</td>
<td>54</td>
<td>102</td>
</tr>
<tr>
<td>/ʌ, o/</td>
<td>83</td>
<td>12</td>
<td>95</td>
</tr>
<tr>
<td>/ɪː, ɪ/</td>
<td>62</td>
<td>32</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 2.6: Summary of ‘principal vowel errors’ in NigE phonemes (Tiffen 1974: 249)

Weighed by auditory judgements of native English assessors, Tiffen reports that the fusion of /iː/ with /ɪ/ and /uː/ with /ʊ/ ‘do not appear to be problems as far as intelligibility in connected speech is concerned’, and notes that the major causes of perception difficulty
‘occur mainly with central and open back vowels’. Inter-ethnically, the test records higher ‘mispronunciation’ ratio for Yoruba speakers than Hausa respondents in /ɜː/ NURSE, /ʌ/ STRUT at 114 to 49 and 85 to 25. His scale yields a higher ratio of deviation for Hausa than Yoruba speakers in /æ/ TRAP, /aː/ BATH and /iː/ FLEECE, and spots the realisation of /ɜː/ NURSE as generally difficult for both ethnics. Other high points include a comparative test for the non-differentiation or distinction between pure tense/lax vowels. While Hausa respondents had a lower frequency of merger of STRUT/NURSE, LOT/THOUGHT, FOOT/GOOSE, DRESS/TRAP and FLEECE/KIT; higher ratio is reported for Yoruba speakers. The absence of contrast in STRUT/THOUGHT and ambiguous realisation of /ɒ/ LOT and /ɜː/ NURSE as well as those marked as less ‘problematic’ for Hausa respondents are most common in the error inventory from the wordlist tokens. These however do not constitute intelligibility failure in the interview data.

Evidently, the works of Nuttall, Afolayan and Tiffens are largely inspired by the Pratorian construct of structural prescriptivism. This is however understandable as most of earliest inquests into NigE patterns were conducted by native speaker researchers – who customarily recommend a stiff compliance with British phonological rules, and as a result dismiss regional stereotypes as deviant or blatant ‘errors’. Tiffens and his contemporaries, for instance, consistently use the terms: errors, mispronunciation, failures, etc in their studies of NigE, and assessed social predictors against the frame of RP rather than the local space of propagation and use. Their works also leave much to cover in the internal mechanism of substrates between competing systems. Apart from Afolayan (1968), patterns of phonological inflections and inter-systemic tensions between the contact languages are hardly explained. Considering the age of these studies, the want of acoustic proof in their analysis may be excused; nevertheless, data elicitation and interpretation processes are basically subjective and in most cases fall short of variationist standards.

Noticeably, there are major differences between the vocalic elements of all Nigerian languages and English. With regard to inventories, none of the local tongues has more than 10 vowels, whereas the SBE has an approximate number of 24 phonemes (depending on the dialect). In fact, what is often held responsible for coalescence and vowel displacement in L2 accents is this disparity. Jibril (1986) detects that the English of Educated Yoruba and Igbo speakers has about 12 vowels in all, whereas their Hausa friends consist of 19 vowels, thus resulting in phonemic coalescence of some central vowels with adjacent items in the cardinal spots.
By these accounts, the resources available for the Northern accent are relatively robust with about 13 vowels of English compared to just 7 in the South. Hausa English system appears more pliant to length and diphthongal distinctions than its Southern counterparts (Table 2.7). Studies which have discussed the triggers of such divergences for this divergence including Awonusi’s (1986) and Banjo’s (1971, 1993) and a few others notionally attribute it to the schooling and teaching models of the earliest schools in both regions during the colonial period (Section 2.1). Jibril believes that certain vowels in Hausa L1 system share allophonic qualities with English, which stands to reason why its educated speakers demonstrate the ease of close approximation with the SBE features. These claims however still await further empirical validation.

In a summary of previous findings by Gut (2004), FLEECE & KIT in Hausa narrowly similar to the native SBE variant, whereas the vowel is realised as [i] or [iː] in the South resulting in a homophonous overlap with FLEECE items. She equally highlights other trajectories including the tendency to freely vary between [ɛ] and [ɛ] for DRESS in typical Southern NigE accent, but also as [ə] or [a] in Hausa English. Echoing similar reports in Eka & Udofot (1996), she notes the singularisation of TRAP and BATH in both Hausa and Southern varieties; and adds that while LOT and THOUGHT have a low back status in Southern accents, they are further lowered and fronted to sound as [a] in mostly North varieties. The FOOT vowel in Hausa English sounds similar to the SBE variant as in /ʊ/, but realised as [u] among Southerners resulting in the non-distinction of full and fool and pull and pool (Simo Bobda 1995, Banjo 1995, Awonusi 2004). Some Igbo speakers may, however, also achieve the pharyngealised variant [u³] of this phoneme (Gut 2004:820). While

<table>
<thead>
<tr>
<th>Educated Hausa English</th>
<th>Educated Southern English</th>
</tr>
</thead>
<tbody>
<tr>
<td>/iː/</td>
<td>/a/</td>
</tr>
<tr>
<td>/i/</td>
<td>/ɛ/</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>/oː/</td>
</tr>
<tr>
<td>/ɛː/</td>
<td>/aː/</td>
</tr>
<tr>
<td>/ʊ/</td>
<td>/a/</td>
</tr>
<tr>
<td>/o/</td>
<td>/i/</td>
</tr>
<tr>
<td>/iː/</td>
<td>/u/</td>
</tr>
</tbody>
</table>

Table 2.7: Monophthongal inventory of Educated Hausa (Northern) and Yoruba/Igbo (Southern) English (Jibril 1986, Jowitt 1991)
Southerners may shorten the BATH vowel to [a], it is marked with length contrast among highly educated northerners (Jibril 1995).

In agreement with Afolayan (1968) and Eka & Udofot (1996), the concurrent lowering and backing of NURSE /ɜ:/ which results in homophonic overlap with [a] has also been attested for Hausa speakers by Jowitt (1991) and Jibril (1995) who both confirm that while most Hausa speakers often pronounce NURSE as BATH, the vowel is orthographically prone to any of [ɜ], [ɔ], [ɛ], [e] or [a] in Yoruba English (Section 2.5.1 for further discussion on this vowel). While other studies appear to limit the non-distinction of DRESS from /eɪ/ FACE in mostly Eastern accents, Jibril (1995) considers a similar possibility for certain Yoruba English variants, and also realisable as [a] in lower lects of Hausa English. Concerning the low back classes, he reports that while STRUT and THOUGHT lack quality and length distinction for southerners, THOUGHT may sound as GOAT [oː] in Hausa English, or diphthongised as MOUTH [au] depending on linguistic contexts. And as summarised in Gut (2004), Jibril agrees with the non-distinction between FOOT and GOOSE for South, and the possibility of pharyngealisation [uˤ] in Igbo English. Interestingly, both vowels are however differentiated in Hausa English (cf. Section 5.2.1)

2.6.1 High, Low Back and Central Vowels

Vowel coalescence and differentiation are some of the most frequently studied sound changes. Their realisation in the speech of individuals as well as their spread through a community raise a number of interesting questions for sociolinguists, dialectologists, phoneticians, and phonologists (Nycz and Hall-Lew 2013, Clark, Watson & Maguire 2013). For example, the English merger phenomenon appears as effectively primordial through its phonological evolution(s). In a diachronic survey by Hickey (2004) on the status of English mergers and near mergers (Table 2.8), he highlights a well-known instance of a previously long mid front /e/ MEET which was later raised to merge with /iː/ MEAT during the Great Vowel Shift, making both sets indistinguishable till date. Most other cases of mergers – phonetic and phonological, i.e. qualitative and quantitative basically follow this paradigm. Hickey’s argument was based on the impracticality of unmerging previously distinct phonemes by subsequent generations of speakers not originally privy to the contrast; given that ‘(i) the merger is not just phonemically, but also phonetically complete and (ii) that language learners are not exposed to varieties of their language in which that merger has not taken place’ (Hickey 2004:1). Mergers in English are generally vocalic as they mostly involve the
collapse of separate vowels into one sound; and can be triggered by largely linguistic or social factors. More often, nonetheless, mergers are preponderantly sensitive to certain phonological contexts (Labov, Ash & Boberg 2006b, Brato 2012, Hofmann 2015).

Studies in sociophonetics, especially those with exploratory goals in dialectology and lectal variations are often concerned with questions on mergers, and the extent to which they illuminate our understanding of occurring and changing trends among speaker groups. Literature abounds on sundry theoretical dimensions of mergers across major established varieties, and the pattern or extent of spread onto emerging ones. This sub-section is thus committed to a recap of this drift across major English varieties and its nuances in key African English systems.

From the Great Vowel Shift to the Second and Third Vowel Shift and recently, the Early Modern, accounts of vowel changes are replete with references to milestones of phonological modification and the structural sway of their residues on English accents around the world. Most cases of mergers in English are often attributed to these shifts, especially for pure vowels in the Early Modern Shifts, i.e., from /i/ lowering to /ɪ/, /e/ to /ɛ/, the raising of /æ/ to /æ/, and the lowering of /u/ to /ʊ/, and /o/ to /ɔ/, and /ɔ/ to /ɒ/ (Schendl & Ritt 2002: 410). For front high and mid vowels, only a few cases of mergers have been attested in both American and British accents. Other than the phenomenal Chain Shifts and the ubiquitous merger of KIT/DRESS before the nasals /m, n, ŋ/ and between FLEECE/KIT before prevocalic /l/ in Southern American accent, high front vowels are mostly retained as distinct by length and quantity in most accents of England, Wales and Southern Ireland (Wells 1982). Beyond the shores of the inner circles, however, myriad cases of contrast loss in the high front and back positions have also been amply reported.

In a study of a lesser-known variety of English by Davey on Gilbraltar English – a territory situated on the Southern edges of the Iberian Peninsula close to the Mediterranean, he corroborates earlier reports of variation by Kellermann (2001) and Lavey (2015), who already note the absence of distinction between FLEECE/KIT, FOOT/GOOSE, LOT/THOUGHT and TRAP/STRUT & START, thus consistent with the trend in most African varieties.
1) The merger of ME /ɛ:/ and /e:/ to /i:/ (meat / meet) (general southern British English)
2) The merger of ME /ai/ and /a/ to /ei/ (tail / tale) (general English)
3) The merger of /ɒ/ and /ɔ/ to /ɑ/ (cot / caught) (forms of American English)
4) The merger of the SQUARE and NURSE lexical sets to the NURSE value (fair / fur) (forms of Ulster English, recent Dublin English)
5) The merger of /uə/ and /ɔ:/ in words like poor and pour (forms of RP)
6) The merger of /v/ and /w/ to [β] (vet / wet) (18c/early 19c southern British English)

1) The merger of /ɔː/ and /ɔr/ to /ɔ:/ (morning / mourning) (most varieties of English)
2) The merger of short vowels before historic /t/ to /ɔː:/ or /əː:/ (tern / turn) (most varieties except perhaps Scottish and Irish English)
3) The merger of /e/ and /æ/ to /et/ before nasals (pen / pin) (south-west Irish English, southern American English)
4) The merger /ei/, /e/ and /ɛ/, often to /e/, before /t/ (Mary / merry / marry) (to varying degrees in various forms of American English)
5) The merger of /ɛ/ and /ʌ/ before /t/ (merry / Murray) (Philadelphia English)

Table 2.8: A list of context independent and dependent merger paradigms in English varieties (Hickey, 2004: 2)

The low back and central vowels, undoubtedly, constitute the famous hotbeds of singularisation already attested universally for most English varieties. Hall–Lew (2010:119) traces the mutation of low back vowels as having ‘been through quite a number of positional and qualitative changes throughout the history of English, resulting today in a complex distribution of lexical sets that variably correspond to different phonemic classes in different areas of the English speaking world’. The Atlas of North American English cites the low-back vowel merger and the allophones of TRAP as the two major pivot conditions for vowel shifts in North American English (Labov, Ash & Boberg 2006). For most British accents and in outer circles, the LOT, THOUGHT and BATH vowels belong to the low back classes. However, members of these sets in the pre-/t/ contexts such as NORTH, FORCE and START are rarely considered in the analyses of mergers (for British and Anglophone African varieties, for instance) due to the acoustic influence of their environment, as well as LOT and PALM sets with central realisations in many American varieties.

The occurrence of low back merger is surprisingly sparse for varieties of British English. Commonest occurrences have rather surfaced more in American dialects where the LOT/THOUGHT sets map comparatively onto COT/CAUGHT (similar to Johnston’s paradigm (Johnston 1997b)). In his account on Scottish varieties, the two classes are considerably merged for urban or young Scots in Glasgow, but clearly distinguished by older
Scots with the CAUGHT realisation apparently lowered to MOUTH (Johnston 1997b: 435-99).

Following a robust empirical tradition set off by Herold (1990) on the distribution of low back mergers in Eastern Pennsylvania, Clopper et al (2005) in their characterisation of six regional American accents, report the ‘partial merger’ of LOT/THOUGHT for the dialects of New England, Mid-Atlantic and Midland but differentiated in Northern and Southern dialects, as well as the effects of sociolinguistic variables in production and perception of mergers in sound changes among San Franciscans (DeCamp 1953, Moonwomon 1992, Labov, Ash & Boberg 2006b and Hall-Lew 2009, 2010). The most interesting of these, in terms of theoretical similarity to the current study is Hall-Lew – who investigates the extent of influence the phonological structure of bilingual speakers have on their contrast realisations or low vowel mergers, and argues ‘that non-English influences on a given variety of English may in fact be necessary for fully understanding the nature of variation in that variety’ (Hall-Lew 2010:126). Her findings thus shows that while the spread of merger is yet to become universal among San Franciscans, low back realisations has active correlation with social factors such as age, hence predicting a complete tilt towards merger in apparent time; and reports a higher correlation of merger score with age among Asian Americans than the European Americans (Hall-Lew 2010: 157).

Also, in a follow-up study to the classic survey by Labov et al. (1997), Irons (2007) in a theroretical confrontation with previous studies on Kentucky English, uncovers the expansion of mergers beyond areas where it was attested to have only been present, i.e., beyond Lexington and Ashland Huntington areas, and suggests that the low-back vowel merger in Kentucky is achieved owing to the off-glide deletion of an already unrounded variants of THOUGHT sets (Irons 2007:162-3).

In contrast to mergers, distinction between previously merged vowels is known as split. Comparatively, ery little of this trend has been described, a consequence of which is, most plausibly, the areal dominance of mergers, i.e. its natural expansion at the expense of distinctions (Labov 1994:313). The paucity of cases regardless, vowel splits remain of equal theoretical and practical concerns in contexts of merger assessment (Nycz & Hall-Lew 2014). For instance, in their test to determine the extent to which some five native Canadian-New-Yorkers have lost their merger of COT/CAUGHT to the historically split ecology of New York observe that – in spite of their substrate residues – a gradience towards split, mostly in the F1 was in progress.
The trajectories of split and merger often appear to bear differently on the West African English varieties from the established inner circle varieties. The FOOT/GOOSE, TRAP/BATH, KIT/FLEECE, and LOT/THOUGHT mergers which have become stereotypical of African accents are still differentiated as separate classes in most British and American varieties. Conversely, apart from Mutonya (2008:446) on young male Zimbabwean English, little of the widespread merger of KIT/DRESS and the raising of TRAP to DRESS due to the chain shift phenomenon (see Hollet 2005, Labov, Ash & Boberg 2006, Boberg 2010, Hofmann 2015) has been reported for the Anglophone varieties (cf. Jibril 1982, Jowitt 1991, Simo Bobda 1995, Awonusi 2004). An interesting account of phonemic transition similar to the Canadians living in New York is a study of five Scottish MPs in Nycz & Hall-Lew (2014). Historically, the speakers’ native variety which has a single low vowel differs from the Standard Southern British English (SBE), whose TRAP had since undergone a split, leading to a more fronted and backer realisation as in TRAP & BATH respectively. Their investigation of the extent to which the MPs had accommodated the in-house accent of the parliament shows similar results to the Canadian-New-Yorkers’, thus describing the movement as a ‘near-split’ achieved more in F2 than F1.

Though there are reports of similar behaviours in some African varieties (Table 2.9), the studies, however, lack relevant variationist perspectives on which questions can possibly be pitched for subsequent investigations. The literature on African English phonology has so far been awash with observations, many of which are difficult to replicate due to grave methodological flaws or utter absence of it (Gut 2004). My commitments in the present study, however, entail a general appraisal of merger instances earlier attested in NigE varieties as well as other West African accents so as to determine their scope of spread and consistency with present data across social and linguistic contexts.

The commonest similarities to the merger phenomenon in West African English accents mainly include the high front, back and central vowels. Though variations may occur on prominent predictors such as ethnicity and level of formal learning in English, speakers tend to share universally consistent patterns of realisation of high vowels (Christopherson 1954, Nutall 1961, Ekong 1978, Awonusi 2004, Mutonya 2008). The lack of distinction between FLEECE and KIT has been consistently reported for educated Southern speakers but
differentiation in Northern accents (Jibril 1982, Jowitt 1991, Simo Bobda 1995, Awonusi 2004). A corresponding pattern applies for FOOT and GOOSE – similarly realised as separate classes by educated Northern speakers (Eka 1985, Odumuh 1987, Awonusi 2004). While a majority of Southern speakers seldom differentiate between the vowels, some appear to favour a more rounded allophone of GOOSE, probably due to the substrate impacts of the L1, especially for speakers with ATR distinctions in high back vowels. The retracted long /u:/ in most Nigerian ATR systems is further lowered on the F1 often resulting in a near fusion with LOT/THOUGHT in an inter-lingual vowel envelope.

There is no consensus as to what factors are responsible, but speakers’ level of education and ethnicity often rank as strongest variables (Bamgbose 1995, Banjo 1995, Udofot 2004, Awonusi 2004). In what Josiah (2009:260) codes as ‘basic’ NigE realisations, GOOSE and FOOT are often non-differentiated in spite of intervening indices such as education and region/ethnicity. In a similar study on Black Kenyan English (BIKE), Hoffman’s (2011:161-162) findings tie in with Schmied’s (2004) on the loss of contrast in high vowel pairs. He observes a ‘very high and peripheral’ realisation and strong merger for FLEECE/KIT [i] and GOOSE/FOOT [u], as well as u-fronting in certain GOOSE tokens – a pattern classed as USE-realisation in this study. An additional mixed effect analysis to equally determine the quantitative dimension of the merger yields no significance, thus suggesting a complete absence of contrast between the high front vowels. In his accounts, though length disparity between the pair of FOOT (with shorter length) and the tokens of GOOSE (two) is noted; no further clarification is indicated as to whether the distinction is merely phonological or purely phonetic in nature (Hoffman 2011:162). His finding however contrasts with Awonusi (2004) who reports a quantitative contrast for the pairs among educated NigE speakers.

The fusion of LOT and THOUGHT including the backed status of STRUT are pervasive in NigE, thus constitute a prominent variation marker. The LOT and THOUGHT contrast is not only absent, STRUT, as a rule, is strongly backed and shifts radically to the low back position (Jowitt 1991, Eka & Udofot 1996, Simo Bobda 1995, Udofot 2004). Curiously, the argument in Ekong (1978) & Awonusi (2004) is that such occurrences are rare in the mainstream educated NigE accent – a claim definitely open to further reviews. On the other hand, Mutonya (2008) in his inter-regional study of African English variations reports a strong culture of STRUT-lowering which results in coalescence with TRAP in typical Kenyan and Ghanaian accents. For younger Ghanaians generally, STRUT however remains unbacked
and excluded from the low back cluster (Mutonya 2008: 445). However, it is not clear in Mutonya’s analysis what other predictors alongside age are responsible for this radical departure.

Some realisations of NURSE do also sound as the NigE variant of STRUT (Jowitt 1991, Simo Bobda 1995) or as START or TRAP in Northern accents (Jowitt 1991). DRESS demonstrates a fair national evenness in quality for most NigE accents, but Odumuh (1987) lists FLEECE and letter as possible variants. Also, a possibility of DRESS-lengthening has been attested in Ekong (1987) and Jibril (1995), but it is not clear what the determinants are. Unlike in some other varieties, distinction between TRAP and BATH sets are completely lacking in NigE. They are usually clustered in the low central position in all varieties and are hardly distinctive by length (Christopherson 1954, Ekong 1987, Banjo 1995, Bamgbose 1995, Adetugbo 2004, Olajide & Olaniyi 2013). In wider contexts, however, TRAP is fronted and raised towards DRESS among young male Zimbabweans (Mutonya 2008:446). Given phonemic stability, the BATH/TRAP vowel is typically contrastive by advancement with much lower F2 in typical British accents (Hall &Nycz 2014).

In addition to instances of coalescence between FACE and DRESS as well as GOAT with the low back group in BIKE, Hoffman (2011:163), based on auditory assessment also uncovers the plausibility of ATR inspired tense/lax dichotomy between the sets, noting that ‘while FACE and CLOTH appear to be tensed (i.e. realised as [e] and [o], respectively) DRESS and GOAT / NORTH are perceptibly ‘laxer’ (i.e. [ɛ] and [ɔ], respectively)’. The fusions, however, in further statistical test show consistency only in height (F1), but a significant difference in F2 values of DRESS than FACE; and CLOTH/FORCE than GOAT/NORTH. Based on these trajectories therefore, he explains the difference in F2 on the +ATR/–ATR system of the speakers’ L1 (Fulop, Kari and Ladefoged, 1998:87). He thus alludes that ‘instead of a tense-lax distinction, it is possible that the perceived difference between these vowel sets is simply due to a [+ATR] – [–ATR] dichotomy: while FACE and CLOTH/FORCE are realised with an advanced tongue root (as [+e] and [+o], respectively, DRESS and GOAT/NORTH are [–ATR] (i.e. [-e] and [-o], respectively’ (Hoffman 2011:163). Doubtful of this logic, however, he cautions against taking it too seriously, hinting on the variance of the F1 results with classic acoustic correlate of an advanced tongue root in most languages which rather is a lowered F1 instead of F2. Also, the correlation between the two systems, i.e. BIKE speaker’s L1 and the English accent would appear spurious as the study
neither shows evidence of the speakers’ L1 phonological features nor plausible manner of substrate transfer.

2.6.2 COMMA/LETTER Lowering

The schwa /ə/ in NigE is one of the displaced vowels from the central positions, but Jibril (1982) and Jowitt (1991) report that, unlike for Southerners, educated Hausa speakers routinely reduce it to a low central allophone of [a]. In RP for instance, the schwa mostly occurs in unstressed syllables by yielding some prosodic force to the adjacent syllable(s). Udofot (1990) notes, however, the rarity of this vowel in Nigerian accents – thus unlikely for speakers to actualise its phonetic gradience in syllabic contexts. Owing to the syllable-timed structure of the NigE prosody, the reduced schwa receives equal articulatory attention as stronger vowels (Gut 2002).

For instance, a wordlist data pooled from 100 educated Yoruba native speakers and a British in Akinjobi (2006) reveals a consistent lack of perceptible prominence between the schwa carrying syllables and adjacent forms (Table 2.10 and Figure 2.4). For a frequency analysis of /ə/, she provides, for elicitation, some 20 suffixed items, with common homogeneous stress-shifting potential. The results show stress retention for all comma tokens in unstressed positions. About 92% respondents demonstrate considerable lowering to BATH/TRAP [a] in grammar for instance. In words like Germanic, 95.7% of strong forms are spotted – in line with earlier assumptions. Also, her mean results for syllable duration interestingly have a higher score than unreduced portions at 141.9ms and 144.4ms respectively.

Based on this finding, LETTER and COMMA are rarely reduced by the educated Yoruba speakers (EYE). Traditionally, a strong syllable in the stem could undergo weakening when suffixes are added. But such reduction is seldom in EYE. Most African languages are assumed syllable-timed in rhythm, whereas their Germanic counterparts are stress-timed (Jowitt 2000, Gut 2005, Udofot 1997, 2011), a view often held on arid empirical justification (Cruttenden 2008). If rythymicality is contingent on the duration of prosodic elements, the above report might also, in the stance of earlier observations, endorse such notion. However, for an actual determination of schwa-tensing in weaker syllables, the need for further instrumental evidence is necessary, taking into account, also, the formant values.
<table>
<thead>
<tr>
<th>Root/ Stem + suffixes</th>
<th>RP pattern</th>
<th>Vowel quality</th>
<th>NE variants</th>
<th>Vowel quality2</th>
</tr>
</thead>
<tbody>
<tr>
<td>pho'netics → pho'neticians</td>
<td>e → ɪ</td>
<td>strong - weak</td>
<td>e → ɛ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'politics → poli'tician</td>
<td>ə → ɔ</td>
<td>**strong - strong</td>
<td>ə → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'comedy → comedian</td>
<td>ɔ → ɔ</td>
<td>strong - weak</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'grammar → gra'mmarian</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>ə → e, a → ɔ</td>
<td>str-str *str-wk</td>
</tr>
<tr>
<td>'Canada → Ca'nadian</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>a → a</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'Colony → Co'lonian</td>
<td>ə → ɔ</td>
<td>strong - weak</td>
<td>o, ɔ → o ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'final → fi'nality</td>
<td>θ → æ</td>
<td>weak - strong</td>
<td>a → a</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'atom → a'tomic</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>a → a</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'tutor → tu'torial</td>
<td>ɔ → ɔ</td>
<td>weak - strong</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'drama → dra'matic</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>a → a</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'strategy → stra'tegic</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>a → a</td>
<td>strong - strong</td>
</tr>
<tr>
<td>e'radicate → eradi'cation</td>
<td>ɪ → æ</td>
<td>**strong - strong</td>
<td>ɛ → ɛ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>ex'hort exhor'tation</td>
<td>ɪ → ɛ</td>
<td>weak - strong</td>
<td>ɛ → ɛ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>e'conomy → eco'nomic</td>
<td>ə → ɔ</td>
<td>weak - strong</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'execute → exe'cution</td>
<td>ɛ → ɛ</td>
<td>strong - weak</td>
<td>ɛ → ɛ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>inferior → inferi'ority</td>
<td>ɔ → ɔ</td>
<td>weak - strong</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'curious → curi'osity</td>
<td>æ → ɔ</td>
<td>weak - strong</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'photograph → photogra'pher</td>
<td>ɔ → ɔ</td>
<td>strong - weak</td>
<td>ɔ, ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>ge'ography → geo'graphic</td>
<td>æ → ɔ</td>
<td>strong - weak</td>
<td>ɔ → ɔ</td>
<td>strong - strong</td>
</tr>
<tr>
<td>'German → Ger'manic</td>
<td>ɔ: → ɔ</td>
<td>strong - weak</td>
<td>a→a, e, a→ɔ</td>
<td>str-str str-wk</td>
</tr>
</tbody>
</table>

276 *24
287 *13

* Appropriate use of weakening
**Appropriate use of strong vowel due the presence of secondary stress

Table 2.10: Frequency table showing the perceptual frequency of ‘appropriate (24 & 13)’ and ‘inappropriate’ (276 & 287) use of schwa in comma /ə/ in unstressed syllables (adapted from Akinjobi 2006:12)

Figure 2.5: Difference in duration between stressed –ne– of phonetics and the unstressed –ne– in phonetician (from Akinjobi 2006:13)
2.6.3 NURSE Lowering, Backing or Fronting

The NURSE vowel is exceptionally diverse in NigE, and in most West African varieties (Simo Bobda 2000:41). The first major account of inter-regional variation of NURSE is Schmied’s (1991a) who in his seminal study on African English varieties notes the backing of /ɜ:/ to /ɒ/ in West African English, lowering to /a/ among East Africans, and fronting to /e/ in South African accents. Instances of backing which often results in coalescence with LOT/THOUGHT and STRUT depending on the accent or fronting towards DRESS as well as lowering are extensive, all of which have been widely attested in NigE literature (Eka 1985, 2004, Odumuh 1987, Jowitt 1991, Adetugbo 2004). The Yoruba English accents, for instance, have been marked as prone to pronouncing the classes of NURSE, STRUT, LOT & THOUGHT as allophones of [ɔ] in or, ur, our orthography and occasionally, first, bird, dirt, third as [fɔst], [bɔd], [dɔt], [tɔd] (Simo Bobda 2000, Gut 2004, Josiah 2011). In a descriptive study of Black Kenyan English, STRUT and NURSE are lowered to cluster with the low central vowels, and remains distinct from on the F1/F2 from low back vowels (Hoffman 2011: 162). Across NigE lects, the extreme lowering of NURSE which invariably results in its non-differentiation from TRAP, especially in term, Bert, earn, learn, firm, birth, myrrh as [tam, bat, lan, bat, fam, ma] for the orthographic er, ear, ir, yrrh (Simo Bobda 2000, Awonusi 2004, Adetugbo 2004). The diagraph ir in girl however, represents an interesting twist as in [gɛl] (Simo Bobda 2000:42). A similar behaviour has been observed for Igbo English where or, ur, our of NURSE words have THOUGHT realisations; and as [ɛ] for er, ear, ir, yrrh in the monosyllabic words. A more radical extension of NURSE/BATH fusion resulting in [wald, ban, mada, fast, bad & pasɜn] for world, burn, murder, first, bad & person is however more indexical of Hausa English (Jowitt 1991).

In terms of social dimensions, the pronunciation of first as [fɔst], [fɛst] or [fast] by Yoruba, Hausa or an Igbo speaker is often critical to ethnic identity. Beyond the national borders, the Cameroonian variety has the graphemes or, ur, our as [ɛ] and [a] in rare variants at the lower bands of the sociolinguistic continuum (Simo Bobda 2000:42). While [ɛ] seems to be the commonest alternants of /ɜ:/ for nearly all lexical incidence of NURSE in Ghanaian English, an African with first as [fɔst] or work as [wɔk] would have either been from Southern Nigeria or Ghana (Mutonya 2008: 445). By the same tokens, the tendency towards NURSE as either TRAP or DRESS is similarly as indexical of East African varieties (Schmied 1991a, Mutonya 2008).
2.7 Diphthongs

Regarding complex vowels, the simplifications of high front, low back and merger of front mid diphthongs or plain reduction to monophthongs are universally typical of NigE accents. In the much earlier study by Schafer (1967) and Afolayan (1968), FACE and GOAT are realised as the half close /eɪ/ and /oʊ/, while /eɪ/ FACE and /eɑ/ SQUARE are ostensibly non-contrastive for Yoruba English. In fact, the pairs of /ɔʊ/ GOAT and /ɑʊ/ MOUTH, /ɑʊ/ PRICE and /eɪ/ FACE in Hausa variety may sound as allophones, with simultaneous lowering of FACE to /e/ DRESS in Hausa variety, and the possibility of coalescence with /ɪɑ/ NEAR (Nuttall 1961, 1965). Despite concurrence with some of Tiffen’s reports, he notes the odds of monophthongal realisation of /ɔɪ/ CHOICE/ as /ɔ/ THOUGHT for some of his Hausa speakers, and the consistent fronting of MOUTH nucleus in Yoruba and Hausa accents. Generally, he marks the /eɪ/ FACE as the only set ‘that presents real barrier to intelligibility for both groups of speakers’ (Tiffen 1974:275).

In most recent studies however, the PRICE and CHOICE vowels have been reported as fully diphthongal for most NigE accents, but with a centring nucleus of /ɑ/ in PRICE and MOUTH in Hausa variety. Typically, all NEAR are realised as /ɪɑ/ (same as SQUARE) or with the epenthetic approximant /j/ to sound as [ija] (See also Eka, 1985; Odumuh 1987, Olajide et al 2013). The CURE words also might be alternated with any of [ʊɑ], [ʊɔ], [oɑ], [ɔː] and [ʊo] (Jibril 1982, Odumuh 1987, Jowitt 1991, Simo Bobda 1995, Eka 2004).

In the wider West Africans contexts, Simo Bobda (2007) surveys the shades of FACE, PRICE, CHOICE, MOUTH, GOAT, NEAR, SQUARE, and CURE vowels (2007:412). According to the account, the FACE vowel is universally monophthongal, and often tends to sound as [e], [ɛ] or [eː], and the PRICE vowel to [a] or [ɛ] in West African varieties. The MOUTH vowel is often pronounced as [a] or [ɑ], whereas CHOICE is much similar to the RP variant as in /ɔɪ/, except for occasional reduction to [ɔ] or [ɛ] for boy, voice, join, spoil, boil, oil most typical of basilects, while GOAT and NEAR could also become [o] and [e], [i], [ɛ] in some varieties (Simo Bobda 2003: 23). CURE has diverse possibilities which include [ɔ], [o] and [u]; and SQUARE is abridged to [ɛ] in Ghanaian and Cameroonian accents, but as the following diphthongs [ɛɛ], [ea], [ie], [ia] in other African varieties.
<table>
<thead>
<tr>
<th>RP</th>
<th>CH</th>
<th>EKN</th>
<th>JI</th>
<th>EK</th>
<th>OD</th>
<th>JO</th>
<th>EK&amp;OD</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>B</td>
<td>V</td>
<td>B</td>
<td>V</td>
<td>B</td>
<td>V</td>
<td>B</td>
</tr>
<tr>
<td>ei</td>
<td>χ</td>
<td>-</td>
<td>χ</td>
<td>ei.e</td>
<td>ei</td>
<td>e;ei</td>
<td>ei</td>
</tr>
<tr>
<td>ai</td>
<td>ai</td>
<td>-</td>
<td>ai</td>
<td>ai</td>
<td>ai</td>
<td>ai</td>
<td>ai</td>
</tr>
<tr>
<td>ei</td>
<td>ei</td>
<td>-</td>
<td>ei</td>
<td>ei</td>
<td>ei</td>
<td>ei</td>
<td>ei</td>
</tr>
<tr>
<td>eo</td>
<td>χ</td>
<td>-</td>
<td>χ</td>
<td>eu</td>
<td>ou</td>
<td>o;au</td>
<td>au</td>
</tr>
<tr>
<td>ao</td>
<td>au</td>
<td>-</td>
<td>au</td>
<td>au</td>
<td>au</td>
<td>au</td>
<td>au</td>
</tr>
<tr>
<td>ë</td>
<td>ia</td>
<td>-</td>
<td>iæ</td>
<td>ia</td>
<td>ɪa</td>
<td>ɪa</td>
<td>ɪa</td>
</tr>
<tr>
<td>ø</td>
<td>ia</td>
<td>øæ</td>
<td>øæ</td>
<td>øæ</td>
<td>øæ</td>
<td>øæ</td>
<td>øæ</td>
</tr>
</tbody>
</table>


Similar to NigE patterns, the high close glide /ɪ/ and the schwa /ə/ are epenthesised with the approximant /j/ in some basilectal varieties, thus yielding [ei]; [aj]; [ej]; [eə] for /eu/; /əu/; /au/; /æu/ (Jibril 1982, Jowitt 1991, Schmied 2004, Hoffmann 2011).

A more empirically focused study, however, is Hoffman’s study of the acrolectal Black Kenyan English (BIKE) using the Lexical Set as paradigm. In separate comparative assessment of traditional complex vowels with the pure phonemes, he reports a case of GOOSE fronting (i.e. /u:/ in /ju:/ contexts – labelled as USE vowel in sets for current study), and shows that the differences between the F-onset and F-offset of GOOSE and those of PRICE set are statistically significant compared to other items such as: BATH, CLOTH, DRESS, FACE, FLEECE, FOOT, FORCE, GOAT, GOOSE (/ju:/ contexts), KIT, NORTH, NURSE, START and STRUT.
Diphthongs | Patterns of monophthongisation
---|---
FACE | >ɛ | >ɛ: | >ɛ
PRICE | >a | >ɛ
CHOICE | >ɔ
MOUTH | >a | >ɔ → ɔ | >a
GOAT | >o | >ɔ
NEAR | >i | >ɛ | >ɛ
SQUARE | >ɛ
CURE | >ɔ | >o | >u

Table 2.12: Patterns of diphthong reduction in African Englishes (Simo Bobda 2007: 414)

2.8 Research Questions

As I have discussed in this chapter, previous accounts on NigE accent have been mainly anchored on meagre or no procedural supports, and particularly in the case of Ebira English – for which there is yet to be evaluation of any sort. What, however, has so far been broadly conjectured is an assumption of correlation between lectal and regional/ethnic differences in the in Nig supersystems. Despite recommendations and proposals towards commitments to ‘minor’ varieties of NigE (Gut 2004, Jolayemi 2006), studies have largely focused on strands comprising the ‘big three’, namely: Hausa, Yoruba & Igbo or on regional classifications such as the Northern and Southern varieties. As also hinted in Jowitt’s (2014:8) critique of Ugorji’s grouping, this contraption has not only been obviously conservative and reductive, but linguistically unfair (Section 1.1).

The procedural deficits of the previous works on NigE regardless, the direction of this study is motivated by their reports of inventories and variations in the system(s). For instance, apart from the lack of consensus on the status of certain vowels for groups in previous accounts, the literature remains largely divergent on plausible inventory of NigE vowels (Josiah 2011). While such goal would be nebulous and over-ambitious – especially given the proneness of accent variations to both internal and extralinguistic factors, investigating the vocalic formation of a sub-system (i.e., Ebira system) within the supersystems (i.e., Yoruba & Hausa systems) would be resourceful. Consequently, this study commits to assessing:
(RQ1). the monophthongal inventory of Ebira English system in the context of previous accounts, i.e., the supersystems of NigE (research question 1), and

(RQ2). the variables (sociolinguistic) responsible for variation in the system (research question 2).

2.9 Conclusion

The questions above derive chiefly from the overall trajectory of reviewed literature in this chapter. However, as earlier hinted, the nature of these reports is one which, unfortunately, impedes clear-cut designs for follow-up studies – in that they were carried out without replicable procedures or analytical methods for follow-up assessments. In terms of social variables in the variety, however, the speakers’ ethnicity and level of education has been consistently held as defining. The vocalic inventories assessed in this chapter variably correspond to these variables. While there is need to read, cautiously, most of these previous findings, this study serves as a more empirical follow-up on them. Thus, the scope of analysis centres on the acoustic measurement of vowels whose realisations are crucial to the phonemic inventories and sociolinguistic categorisations in the NigE system. The following chapters details how this is done for the EEng vowels. Generally, I bring to bear – as much as possible, the Labovian design, involving the acoustic phonetic methods through data collection, speech digitisation, formant measurement and extraction, and statistical assessment. The goal, therefore, is to present an imposing account of the vocalic stratum for this variety (RQ1), as well as initiate a more convincing paradigm for explaining sociolinguistic dimensions in other NigE varieties (RQ2).
3 Research Design

3.1 Speakers

In this section, I outline some demographic considerations for recruiting participants for the study. Due to the methodological inadequacies in prior accounts of NigE phonemes, speakers’ stratification was based on standard variationist traditions. As a result, it was fairly problematic to hit the ground running with a set of fine-grained hypotheses or speaker grouping (i.e., a data structure clearly anchored on pre-set variables or research questions). This difficulty, notwithstanding, theoretical regularities abound with regard to correlation between speakers’ accents and factors including: age, gender and socio-economic status, etc (Di Paolo & Yaeger-Dror 2011: 9).

With respect to the nexus between the social class and linguistic structures, the Nigeria’s situation chiefly contrasts most of other regions where strong correlation is often found between the two. Historically, the most significant factors often have little or nothing to do with speakers’ socio-economic status. In fact, some studies have found the level of education as rarely defining (Fakoya 2004, Jowitt 2015, Oladipupo 2016). This trend, however, is least remarkable. Bourdieu and Boltanski (1975), in their bid to disambiguate the complex relationships between social class and language use provide a construct now known as the ‘linguistic market’. They theorise that the major indices of social positioning – such as income weight, educational background, occupation never really ‘distinguish speakers within the same social class whose access to and use of the standard language in a community were quite different’ (cited in Wagner 2012: 375). It is thus possible to find, among exclusive professionals or common practitioners, those for whom faithfulness to certain linguistic norms is less important and those for whom it’s essential (Chambers 2003).

The trend in most well-studied varieties is to conjecture a hunch which the researcher sets out to test against previous theoretical assumptions for a group of speakers. This is important in a number of ways, most especially for the design of this study, as it provides a defined outlook, and mitigates the scope of diagnostic enquiries prefacing the main thesis. For instance, existing theories on a variable can instruct the direction of follow-up studies, or even inspire the brand of procedures (see Hofmann 2015). In the absence of such incentives for the
present study, therefore, the task of speakers’ grouping, data elicitation and the modelling of a linguistic protocol were largely done in the dark – albeit inspired by mainstream conventions. Thus, grouping of cohorts into age, gender, education and the number of L1 (Nigerian languages spoken apart from Ebira) of the speakers were considered the key determinants.

3.2 Sampling

The majority of the decisions I made regarding sampling had little leaning on the literature on NigE – hence my reliance on the basics of the variationist models. Often in sociolinguistic enquiries, either random or judgement sampling (also quota sampling) is considered sufficient (Wardhaugh & Fuller 2015: 159). Owing to the usual constraints in sociophonetic assessments, the judgement sampling is mostly suitable (Chambers 1995: 38–41, Milroy & Gordon 2003: 26, Di Paolo & Yaeger Dror 2011: 13, Hofmann 2015: 133). It involves the stratification of speakers according to some pre-set criteria: age, education, gender, social class, etc; with the aim of filling the corresponding cells with respondents on the field. Accordingly, I ensured that the population entirely consist of fluent Ebira English speakers, whose L1 is Ebira, and had lived for the most part in Ebiraland till the time of data collection.

For expediency, my initial cohort into each quota was done according to age and gender, with education and job status presumably controlled for. And in order reduce the preponderance of poor eloquence in English and also achieve fair representation of what has come to be known as the ‘Educated Nigerian English’ (Brosnahan 1958, Banjo 1971, Bamgbosse 1982, Udofot 2003, Gut 2004), only speakers who already had a tertiary school degree or in the process of bagging one were included. This way, a homogenous level of literacy for all respondents was assumed; but as I will explain shortly, things aren’t always as clear-cut. I noticed just after my fieldwork – that majority of the speakers were split between two socially sensitive academic honours (BSc & HND) in the Nigerian contexts, thereby necessitating the creation of additional sub-set groups for education degree. My intention was to fairly represent all categories across genders, but this could not be strictly achieved due uneven distribution across the levels (Table 3.1)

As explained in Section 3.2.1 below, the quota was drawn in view of possible interaction between age and job status (the equivalent of social status) for EEng speakers. Based on pre-study survey, social ranking in Nigeria is hardly ever as unambiguous as in most other climes whose English varieties have been socially assessed. Most of my older speakers were civil servants who work as teachers or as admin personnel in academic institutions. Only
2 of my older speakers (1 nurse and 1 businessman) were not in this category. None of them was at the time unemployed. The younger speakers, however, were predominantly students or fresh graduates from the state universities and polytechnics. Only one speaker – evidently of the younger age cohort – had just been employed at the time of data collection. I had expected, therefore, this interaction to compensate for the exclusion of job status at pre-field stage, leaving age and gender as core social predictors. As the VIF test for collinearity would later reveal, the combination of job status and age yielded a high collinear effect, hence excluded from the main analysis in Chapter 5.

My data collection commenced on the 2nd May 2014 through the early days of June 2014. Prior to this period, I had four major categories, consisting 6 younger male, 6 older female, 6 younger female, and 6 older female). My goal was to recruit speakers into these groups, for evenness and statistical robustness (Bailey & Dyer 1992, Lawrence 1990). In addition to being an insider, I used the friend-of-a-friend approach (Milroy 1980:47, Milroy & Gordon 2003, Schilling 2013). This method proved very aiding, particularly with the older speakers. Earlier before the interviews, I visited a number of high schools and offices in Okene (the capital city of Ebiraland) to discuss my intentions with prospective respondents and booked appointments with those who were disposed to being recorded. Despite plain declaration of research intent, some individuals either refused to give me a ‘yes’ at the first visit or bluntly declined on concerns about the purpose of the interview. The good news, however, was that respondents who later grew keen on my research happily introduced me to their colleagues and friends within their workplace or neighbourhood. After largely unsuccessful efforts to recruit enough young speakers living in Okene metropolis, I visited a neighbouring polytechnic city, Lokoja – where I was hosted by a college lecturer for 4 days. His involvement turned out a remarkably big boon in a number of ways! Not only did he and his colleagues agree to be interviewed, they also helped recruit students of the institution who also gladly participated. Another subtly aiding factor was the simultaneous recording of wordlist tokens from the older speakers for the Ebira L1 data. Somehow, this fanned their sense of patriotism and inclusion, hence evoked the much needed partisanship for the process.

By the end of my field work, I had recorded 28 speakers in the indigenous districts of Ebiraland, and 2 other speakers at Akungba-Akoko in Ondo State (Table 3.1 and Table 3.2). The inclusion of the 2 speakers recorded in a separate location speakers considered on their fulfilment of pre-set criteria for selecting respondents. Of the 30 speakers, 3 (FUY2 & FUO6) were excluded for notably bad or incomplete speech samples. All speakers were recorded
individually so as to avoid accidental recording – as well as other technical drawbacks; and
the need for clear procedures that could mentor follow-up studies (Hofmann 2015: 135).

<table>
<thead>
<tr>
<th>speakers</th>
<th>gender</th>
<th>age.cat</th>
<th>edu.deg</th>
<th>job.status</th>
<th>additional Nigerian language</th>
<th>age</th>
<th>age@expo</th>
<th>yrs.of.expo</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUO1</td>
<td>MALE</td>
<td>OLDR</td>
<td>HND</td>
<td>Cv. Srvt</td>
<td>Hau. &amp; Yor.</td>
<td>56</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>MUO2</td>
<td>MALE</td>
<td>OLDR</td>
<td>MSc</td>
<td>Cv. Srvt</td>
<td>Yoruba</td>
<td>54</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>MUO3</td>
<td>MALE</td>
<td>OLDR</td>
<td>MSc</td>
<td>Cv. Srvt</td>
<td>Yoruba</td>
<td>35</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>MUO4</td>
<td>MALE</td>
<td>OLDR</td>
<td>BSc</td>
<td>Cv. Srvt</td>
<td>Yoruba</td>
<td>54</td>
<td>10</td>
<td>44</td>
</tr>
<tr>
<td>MUO5</td>
<td>MALE</td>
<td>OLDR</td>
<td>BSc</td>
<td>Cv. Srvt</td>
<td>NA</td>
<td>38</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>MUO6</td>
<td>MALE</td>
<td>OLDR</td>
<td>BSc</td>
<td>Teacher</td>
<td>NA</td>
<td>48</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>MUO7</td>
<td>MALE</td>
<td>OLDR</td>
<td>NCE</td>
<td>Teacher</td>
<td>NA</td>
<td>34</td>
<td>3</td>
<td>31</td>
</tr>
<tr>
<td>MUO8</td>
<td>MALE</td>
<td>OLDR</td>
<td>PGD</td>
<td>Bsman</td>
<td>NA</td>
<td>44</td>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>FUO1</td>
<td>FEM</td>
<td>OLDR</td>
<td>BSc</td>
<td>Cv. Srvt</td>
<td>NA</td>
<td>41</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>FUO2</td>
<td>FEM</td>
<td>OLDR</td>
<td>HND</td>
<td>Cv. Srvt</td>
<td>NA</td>
<td>35</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>FUO3</td>
<td>FEM</td>
<td>OLDR</td>
<td>HND</td>
<td>Teacher</td>
<td>NA</td>
<td>52</td>
<td>8</td>
<td>44</td>
</tr>
<tr>
<td>FUO4</td>
<td>FEM</td>
<td>OLDR</td>
<td>NCE</td>
<td>Teacher</td>
<td>Yoruba</td>
<td>31</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>FUO5</td>
<td>FEM</td>
<td>OLDR</td>
<td>BSc</td>
<td>Teacher</td>
<td>Yoruba</td>
<td>56</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td>FUO7</td>
<td>FEM</td>
<td>OLDR</td>
<td>BSc</td>
<td>Cv. Srvt</td>
<td>Yoruba</td>
<td>38</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>MUY1</td>
<td>MALE</td>
<td>YNGR</td>
<td>BSc</td>
<td>Student</td>
<td>NA</td>
<td>24</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>MUY2</td>
<td>MALE</td>
<td>YNGR</td>
<td>BSc</td>
<td>Student</td>
<td>NA</td>
<td>24</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>MUY3</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Cv.Srvt</td>
<td>Yoruba</td>
<td>28</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>MUY4</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>21</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>MUY5</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>19</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>MUY6</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>26</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>MUY7</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>23</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>MUY8</td>
<td>MALE</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>26</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>FUY1</td>
<td>FEM</td>
<td>YNGR</td>
<td>BSc</td>
<td>Student</td>
<td>NA</td>
<td>25</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>FUY3</td>
<td>FEM</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>24</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>FUY4</td>
<td>FEM</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>23</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>FUY5</td>
<td>FEM</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>20</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>FUY6</td>
<td>FEM</td>
<td>YNGR</td>
<td>HND</td>
<td>Student</td>
<td>NA</td>
<td>19</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>FUY7</td>
<td>FEM</td>
<td>YNGR</td>
<td>BSc</td>
<td>Student</td>
<td>Yoruba</td>
<td>22</td>
<td>4</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.1: Speaker name, gender, age-category, education degree, other Nigerian languages, spoken, chronological age, age of exposure to English and years of exposure. Key Abbreviations: HND (Higher National Diploma), NCE (National Certificate in Education), MSc (Master in Science), BSc (Bachelor in Science), PGD (Postgraduate Diploma), MUO (Male University Older), FUO (Female University Older), MUY (Male University Younger), FUY (Female University Younger), OLDR (Older), YNGR (Younger)
Table 3.2: Distribution of respondents according to age and sex (= 28)

<table>
<thead>
<tr>
<th>age group</th>
<th>male</th>
<th>female</th>
<th>total</th>
<th>min. age</th>
<th>max. age</th>
<th>mean age</th>
<th>STDEV. Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger</td>
<td>8</td>
<td>6</td>
<td>14</td>
<td>19</td>
<td>28</td>
<td>23.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Older</td>
<td>8</td>
<td>6</td>
<td>14</td>
<td>35</td>
<td>56</td>
<td>43.6</td>
<td>8.8</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>12</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.1 Age

Despite widespread infrequency of correlation between human biological growth and social development, studies in sociophonetic variation overwhelmingly depend on chronological age for grouping speakers into cohorts (Eckert 1998:154). In this study, age was assigned a prime status, given its structuring force on the overall process of linguistic evolution, variations and change (cf. Labov 1972, 1994, Clermont and Cedergren 1979, Trudgill 1988, Eckert 2000, Sankoff 2005, Labov 2006, Chambers 2009, Wagner 2012, Brato 2012, Hofmann 2015). One of the most prominent theories is, perhaps, the Labovian construct of ‘age-grading’ – a model originally deployed in anthropological studies (Hockett 1950:453). It sprang on the evidence of steady correlation between age and nuclei height in his analysis of /ay/ and /aw/ realisation by speakers in Martha’s Vineyard (Labov 1963). The diphthongs’ nuclei were increasingly higher with each younger age cohort which suggests that the older speakers might have, at some point in their early life, favoured the same pattern as their younger counterparts, but gradually commenced the nuclei lowering as their age grew. Thus, age categorisation (Table 3.3) can simultaneously reflect changes in speakers’ behaviour as they move through life stages – age-grading or communal variations in the course of time – generational change (Eckert 1996:150, Di Paolo & Yaeger-Dror 2011:9, Wagner 2012:371). Unfortunately, the choice of age models implementable with my speakers’ types and data was evidently hindered. As an insider, I had anticipated a weak correlation between factors such as the: real time age of speakers, age of exposure to L2, years of exposure to L2 and linguistic behaviour. These variables were recorded, nevertheless, and tested for possible correlation, but none showed any effect, especially in differentiating supposedly merged vowel classes (Section 5.2). For the younger age cohort, those between the ages of 15 – 18 were excluded. Studies have shown speakers in this category as often pliant to unstable peer norms – that are rarely
carried on through the following years when they later commence formal interaction with the larger communities, especially at the tertiary institutions (Eckert 1998, Hofmann 2015). The official entry year into Nigerian tertiary institution is 18 (though it is possible to find undergraduates of a lesser age); and since my aim was to control for education by including only speakers in tertiary schools or those already with a tertiary degree, it would be much harder to recruit respondents from this rank. The cohort also represents the Crèche/Nursery School age, mostly those who were never taught at school with any Nigerian native language, i.e., taught only in English throughout elementary classes. Across both gender, my younger speakers were between the ages of 19 – 25. Two speakers, MUY3 & MUY6, 28 and 26 years respectively, as at the time of interview, were however included in this cohort, given their in-group similarity with other speakers in the younger age cohort.

<table>
<thead>
<tr>
<th></th>
<th>Individual</th>
<th>Community</th>
<th>Synchronic Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stability</td>
<td>Stable</td>
<td>Stable</td>
</tr>
<tr>
<td>2</td>
<td>Age-grading</td>
<td>Unstable</td>
<td>Stable</td>
</tr>
<tr>
<td>3</td>
<td>Generation</td>
<td>Stable</td>
<td>Unstable</td>
</tr>
<tr>
<td>4</td>
<td>Communal change</td>
<td>Unstable</td>
<td>Unstable</td>
</tr>
</tbody>
</table>

Table 3.3: Trend of linguistic change in the individual and the community (from Sankoff 2005:1004)

For older speakers, my curiosity was keener on the pattern of behaviour after retirement which usually begins at 65 or 67, hence my desire to include them. At first, I was motivated by the paucity of studies on this age group in the literature (Eckert 1996:165, Chambers 2009:179, Hofmann 2015:119). These age cohort are often assumed prone to loosen up their conservative speech culture as they gradually disengage from the ‘marketplace’ (Labov 1972, Keith 1980). In NigE contexts however – especially for EEng, very few speakers in this age bracket had the education criterion (a tertiary institution degree) or speak English fluently among the population, or would be very difficult recruit, if there were. Instead, I choose speakers who were through the middle of their career age up to the eve of retirement (35 – 55). This age-bracket is usually linked to some level of linguistic conservatism – inspired and sustained by workplace pressures and institutional social networks (Sankoff & Laberg 1978, Milroy 1980, Edwards 1992, Eckert 1996). Again, two speakers, MUO1 & FUO5, both 56, were also included. Among the speakers, there was a
strong binary split between those who were civil servants or students, thus suggestive of strong inter-level interaction between job status and age.

3.2.2 Gender

Gender portrayal in phonological variation first featured in Labov’s study on Martha’s Vineyard in which the cross-tabulation of speakers’ sex, age and socio-economic status with phonemic variables is significant (Labov 1966; Figure 3.1). Follow-up studies on the linguistic differentiation of men and women have mainly evolved into two distinct principles (cf. Labov 1990:205–220). That in stable sociolinguistic stratification, men use a higher frequency of nonstandard forms than women; but the other way round in cases of change imposed by the upper social caste – in New York City (Labov 1966), Detroit (Wolfram 1969), Glasgow (Macaulay 1977), Norwich (Trudgill 1974), Belfast (Milroy 1980) – and that in change from below, men lead in advance (Labov 1996, Trudgill 1974 & Eckert 1989). By these assumptions, women's language is thus reflective of ‘conservatism, prestige consciousness, upward mobility, insecurity, deference, nurturance, emotional expressivity,

![Figure 3.1: Style shifting of (oh) by male and female speakers across three socioeconomic groups (from Labov 1990: 224)](image-url)
connectedness, sensitivity to others, solidarity; while men's language mostly conveys toughness, lack of affect, competitiveness, independence, competence, hierarchy and control (Eckert & McConnell-Ginet 1992:90). Some of these studies arrogate the power of fluency and overt standard linguistic norm the women, in so far they depend more on their language use for social standing. In the sociolinguistic delineation of speakers, however, sexual conceptualisation ought to be more of gender than biological taxonomy (Labov 1990:221; Wodak & Benke 1997:128). Also, several other confounding variables would require separate weighing in the analysis so as to avoid exaggerated outcomes that are underlyingly due to key extraneous factors. Following a much feasible approach explained in Labov (1990: 221-2), speakers were grouped according to the traditional male and female categories, albeit cross-tabulated with other interacting variables.

Up till the late 1970s in Nigeria, especially in Ebiraland, females’ access to literacy was constrained by culture. Seen as potential wives whose duties would solely involve housekeeping and baby-making, their social engagements were largely dim. Even those who managed to gain some form of early education alongside their male peers grappled with such bias that kept them out of stage. One of my older female speakers took time to narrate the severe challenges she had to deal with as a young pupil up to the university level. Her uncle had advised her father never to afford her the grace of formal training – because she was just ‘a lady’ after all. But the privilege, she recalled, was warmly extended to all her brothers. By the time she eventually wriggled her way into a school at a later age, she told of how often her male peers would cast jeers at her; and the fact that even her teachers thought very little of her prospects too. This unsettling trend, however seemed to have been drastically reversed in a little over the last two decades, and there has been a sustained drive towards gender equality in general. Both genders appear to now have equal access to education. Amid other variables, the gender taxonomy in my data was designed to reflect this historical development (see Table 3.2).

3.2.3 Education

Among the known social indices so far considered on NigE accents, speakers’ level of formal or Western education has been the most dominant. Often corroborated with ethnicity, education has appeared in nearly all variationist attempts to classify the accents of NigE (cf. Christopherson, 1954, Ekong 1978, Jibril 1982, Eka 1985, Odumuh 1987, Jowitt 1991,
The trend is thus very apt and historically consistent with the speakers’ linguistic ecology. Over the years, English has grown to become more powerful than the native language – and for many Nigerians – is daily relied on for a range of communication needs. It is also the principal (more or less the only) language of formal training, religion, commerce, journalism, domestic interactions and emotional bonding (Section 2.1). Since the most viable access to English proficiency from the earliest periods was Western education, the temptation to privilege it above other factors is plausibly strong.

From the outset, education was controlled for, ensuring that all speakers to be interviewed would have either bagged a university degree or in the process of getting one. I intended recruiting all respondents on this criterion in order to ensure that education was controlled for (Brosnahan 1958, Banjo 1971, Bamgbose 1982, Udofot 2003, Gut 2004). As pointed in Section 3.1, situations on the field are not always as intended at in pre-field design. Of the overall number of respondents, 11 have university degrees; 14 have Higher National Diploma; 2 have the National Certificate in Education (NCE); 2 have Masters in Science (MSc); and 1 with a Postgraduate Diploma (PGD). In view of the stealth social ranking of these qualifications in Nigeria (though rarely noted officially), corresponding levels were created for further representations: NCE was merged with HND; PGD with MSc; and BSc as separate category. These subsets have, in prior studies, been often subsumed under ‘educated speakers’ and my intent was to investigate possible effects (if any) between these psychosocial dichotomies and vocalic differentiations.

### 3.3 Sociolinguistics Questionnaires and Interviews

‘...the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation’ (Labov 1972:209)

For the recording of speakers’ information, i.e., meta-data, written questionnaires were handed to each respondent, prior to the interviews. The synergy of self-written bio-data and the accompanying analysis of acoustic data can be highly resourceful in generating social facts about speakers, and setting clear parameters for stratification of key variables (Dollinger

---

13 Higher National Diploma (HND) is obtained from Polytechnics, while Bachelor of Science (BSc) or Bachelor of Arts (BA) is obtained from Universities

14 National Certificate in Education (NCE) is awarded by Colleges of Education as professional certificate to intending high school and elementary school teachers
To keep this as brief and exciting as possible, questions were restricted to one page, and designed mainly to gather the most useful information. The questionnaires were in 4 subsections. Questions about age, gender, city/neighbourhood, place of birth and childhood were in Sections 1 & 2. Thus, name of the districts, city of birth, and travel history from Ebiraland since childhood were asked. Section 3 featured questions on speakers’ academic qualifications and professional career, such as: the layers of formal trainings, highest qualification obtained, career history, and current occupation/job title. The last Section (4) focused on language skills: the onset year of English learning, the extent of use in domestic contexts and if/how many Nigerian language(s) spoken other than the L1.

Getting a corpus of natural speech data is very crucial for sociophonetic analyses. Thus, the goal is typically towards obtaining a predominantly normal speech from the respondents, bearing in mind that speakers tend to maintain their way of speaking when they less self-conscious, and erratically when made self-conscious (Di Paolo & Yaeger-Dror 2011:9). In order to source the ‘purest’ form of speech samples from speakers, and the feeling that respondents do modify their accent to fit into different speech styles, a certain order is deemed ideal, for instance – recording the speakers’ casual speech style prior to more formal way of speaking: the reading passage and the word list (Labov 1972 Trudgill 1974, Patrick 2016). A major benefit of such ordering would be to overcome the ‘Observer’s Paradox’ or interviewer’s effect. Labov believes that respondents can be made to not recall the unnaturalness of the interview situation by getting them to recount their best street games, life-threatening encounter, or spurred on to retell emotional stories that engages their memory at the expense of attention to their manner of speech (Labov 1972: 92). It’s also reasoned that the analyst faces the risk of being ‘found out’, especially if the citation is skewed towards certain minimal pairs or some item historically unique or interesting in the variety.

Examined in another way, evidence from major studies has shown that respondents rarely reflect such dichotomies, and that if they do at all, the order would have little or no effect on the data (Milroy & Gordon 2003:50, Tagliamonte 2006:38). In similar re-evaluations, researchers have observed that most speakers either continue to drift on different styles within and between the interviews; or become threatened at emotion-laden questions that may trigger the discussion of life’s events they would rather keep personal (Eckert 2000, Mendoza-Denton 2008, Schillings 2013:104). Hofmann, during his fieldwork in Newfoundland – for instance, observes that his interviewees’ reaction to the stylistic range of data he used was more determined by the level of interpersonal relationship with them than
sequence of elicitation (Hofmann 2015:120). During my time on the field, the module: ‘Did you ever get blamed [punished] for something you didn’t do?’ (Labov 1973) nearly failed with two of my respondents. One of them had dissolved into tears in the course narrating her experience, whereas a speaker simply declined discussing hers.

As far as field interviews are concerned, things can be kept as flexible as possible, depending on what the situation or context demands, and there should hardly be a straitjacket routine. For instance, my shared ethnicity and socio-cultural experiences with all respondents was a helpful tool of access to the speakers. Prior to my arrival in Okene, I contacted my sister who had been residing in the Local Government Area (LGA), and provided her with details of my study, thus enabling her to contact a few acquaintances who might fit into the demography and willing to participate. She would later double as my host and ad-hoc ‘research assistant’ throughout my fieldwork in Ebiraland. Culturally, every Ebira-born is seen as a family member by kinsmen. Thus, many of my respondents had little issues with blending into talking with one of their own! Apart from the questionnaire I gave them to fill, the Informed Consent certificate and the short letter of introduction handed each speaker to read before the interview, I also ensured chit-chats with them before recordings. Since style was a major variable, I began all sessions with the most formal to informal styles, i.e., word list/citation, the reading passage and the conversation styles.

3.3.1 The Word List Citation

Basically, the citation style comprised a list of words written for speakers to read – some of which embedded phonological co-texts of interest. Though less preferred for testing natural speaking, the word list is crucial in a number of ways. With the citation mode, the predisposition for the speaker to over- or under-report usages is most likely, especially if, by some means, they get egged on wrongly by the researcher. More than other styles possibly would, it provides some sort of levelling paradigm across speakers and phonological contexts. And for vowels flanked by certain co-textual consonants, it guarantees the extraction of uniformly representative tokens that are of key interest for the study. With citations, variable types and contexts can be managed, and the length of attention to each item on the list can equally be controlled (Kerswill 1994, Gordon 2001). A further argument is that, in most varieties, vowels of interest are often below speakers’ level of conscious awareness, thus, can
be safely hidden in the word list items to supplement other styles (Di Paolo & Yaeger-Dror 2011:4).

At kick-off, I originally included in the study, as supplement, some consonants, such as: /l/ vocalisation, t–affrication, and th–alveolarisation. I had intended to incorporate an impressionistic description of these, in addition to the crux of vocalic analyses and variations among the speakers. Thus, I was literally aiming at two birds with a sling. In drawing the list, I followed Wells’ paradigm which includes variants for the British and the American accents (Wells 1982). Considering the historical antecedents of English in Nigeria (Section 2.2), it was plausible to adapt the sets after the British accents; but this would foreclose checking for possible effects of late American imperial influence on the variety. Hence, corresponding items including NORTH, FORCE & PALM classes were added, based on which further tokens were extracted for analysis (Wells 1982).

My word list had 23 disyllabic and 114 monosyllabic items, making a total of 137 words. For each set, at least four items were included. The list consists of no entry beyond two syllables, as pronouncing such might be difficult for some respondents. Also, the inclusion of words with fewer syllables would, to some extent, mitigate the predominant effects of undershoot, as well as ensure a comparative base-line for the vowels (Thomas 2011:174). In phonological environments, 27 items occurred in preceding t-context, 13 before /l/, and 17 sandwiched in either pre- or post-dental fricative. To test for a range of distinct but reportedly merged vowels in the variety, corresponding minimal or similar pairs were disguised within the list. For example, the following were included for KIT/FLEECE: hit/heat, ship/sheep, and filth/field; for GOOSE/FOOT: shoot/choose and tooth/put; LOT/THOUGHT: caught/pot and sought/sock, as well as for BATH/TRAP (see Appendix A.1). Prior to preparing the slides, all items were randomised, ensuring that the words with very similar phonological structures were not near to each other. The slide-show was set in every 2 seconds, with each item written in Calibri/115 font on a 14” display Laptop computer. (Di Paolo & Yaeger-Dror 2011:15). This way, the interviewer could only read a token from a slide at a time, so preventing list intonation and the interviewees neither from looking ahead nor behind. Since the task commenced with the word list, effort was made to ensure that nothing was ‘given out’, especially the minimal pairs. An early discovery of interesting vowels/items could summarily impact the process negatively.
3.3.2 The Reading Passage

The structured narrative, also known as the reading passage often comes next on the formal hierarchy of styles. Unlike the wordlist, the reading passage ranks less in structuredness, and constrains the insertion of minimal pairs for specific variables. As a guided narrative, the reading passage subtly turns the speakers’ attention away from isolated (as in word list) to a storyline strewn on amusing thoughts. It is thus less formal than the word list and less informal than the conversation style (see 3.3.3 below). The inclusion of a reading passage in the interview tasks has proved crucially supplementary, particularly in predicting patterns of stylistic variation (Schilling 2013:98).

![Figure 3.2](image)

Figure 3.2: r-realisation in New York City by socioeconomic classes and speech styles (from Labov 1966:240)

For example, Labov’s analyses of NYC English (Figure 3.2) and Philadelphians’ shows that the gradient progression of the speakers’ styles as the interview moves through the casual, reading passage and the conversation styles. He notes that the speakers, mainly of
different socioeconomic extractions, gradually raised their adherence to formality in the case of post-vocalic \( r \)-pronunciation at the expense of stigmatised variants from the most formal to the informal styles (Labov 1966). From his test of correlation between rhoticisation (an originally Upper Middle Class feature) and socioeconomic status in New York City, Labov uncovers, among other phenomena, an instance of hypercorrection among the lower class – who consciously accommodate, even out-perform the Upper Middle Class in the more formal styles, but slip back to \( r \)-lessness in the casual speech. This finding, to a certain extent, corroborates some observation I made during my interview sessions. Many of the speakers were exposed to what now is referred to as ‘bookish English’ at school, or had become weaned on rote-drills, hence paid more attention to vowel differentiation (especially between minimal pairs) in word list tokens than in less formal styles (Awonusi 1986:557).

Equally imperative is the choice of reading passage. Regrettably, some of the famous narratives often used, including *Arthur the Rat* and *The North Wind and the Sun* could barely cater for the stretch of the sounds under investigation (Labov 2006: 60, Hofmann 2015: 126). For my study, *The North Wind and the Sun* had clear shortcomings. Though it richly represents the monophthongs of English, tokens of /ʒ/, medial and initial /z/, initial /θ/, word-final consonant clusters ending in /s/ or /z/, and diphthongs /ɔɪ/ and /eə/ could not be extracted from the passage. Since my analysis would involve the descriptive representation of the vowel system, using the passage would also complicate the assessment of certain diphthongs missing in the passages. A rather more inventive passage, especially for sociophonetic purposes is *Comma Gets a Cure* (Honorof, McCullough & Somerville 2000). Unfortunately, also, I could not use it for my interview due to the fact that it has much longer (375 words) than *The Boy who Cried Wolf* (221 words). Though the former offers a healthy range of vowel distributions for all the Lexical Sets, the benefits of *Comma Gets a Cure* would be lost, especially given the fact that non-native English speakers might be put off by its length, or have general difficulty within the flow of the narrative. Generally, it was also likely that majority of them would have hard time relating its predominantly ‘foreign’ characters and thematic rendering to their local milieu – therefore failing in naturalness.

*The Boy who Cried Wolf* has also a fair distribution of vowels (Table 3.4). Barring phonemic distributions according to different accents of English (Wells 1982), the text has at least, three clear instances each for all English monophthongs, all of which are measurable in healthy neighbouring environments (Deterding 2006:193). There are also relevant items for complex vowels of interest such as CHOICE /ɔɪ/, SQUARE /eə/; and minimal pairs for
differentiation effect between: fool & full for GOOSE /u:/ & FOOT /ʊ/, feast & fist for FLEECE /i:/ & KIT /ɪ/, short & shot for THOUGHT /ɔ:/ & LOT /ɒ/. For vowel occurrence in co-textual environments, /θ/ occurs in thought, threaten and third, /ʃ/ occurs in pleasure and usual, initial and /z/ in zoo, raising, cousins and exactly. The dark variant of /l/ occurs in final position after full & fool, and as part of a final cluster in wolf, fields and himself; as well as syllabic in little & successful, among others (Deterding 2006:193). In addition, The Boy Who Cried Wolf is set with a culturally neutral ambience and a textual cadence that can be easily followed. In fact, two of my respondents were noticeably amused by the storyline in the course of reading, thus showing that they were more conscious of the narrative rather their way of reading it.

<table>
<thead>
<tr>
<th>KIT</th>
<th>FLEECE</th>
<th>DRESS</th>
<th>TRAP</th>
<th>LOT</th>
<th>STRUT</th>
<th>FOOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>village(2x)</td>
<td>fields</td>
<td>shepherd</td>
<td>ran</td>
<td>flocks</td>
<td>company</td>
<td>foot</td>
</tr>
<tr>
<td>villagers(2x)</td>
<td>sheep</td>
<td>get</td>
<td>exactly</td>
<td>shot</td>
<td>rushed</td>
<td>good</td>
</tr>
<tr>
<td>fist</td>
<td>feast</td>
<td>pleasure</td>
<td>actually</td>
<td>bother</td>
<td>cousins</td>
<td>full</td>
</tr>
<tr>
<td>chicken</td>
<td>even</td>
<td>successful</td>
<td>had</td>
<td>just</td>
<td>looking</td>
<td></td>
</tr>
<tr>
<td>trick</td>
<td>did</td>
<td>threaten</td>
<td>began</td>
<td></td>
<td>duck</td>
<td></td>
</tr>
<tr>
<td>convinced</td>
<td>little</td>
<td></td>
<td></td>
<td></td>
<td>overcoming</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NURSE</th>
<th>MOUTH</th>
<th>FACE</th>
<th>GOAT</th>
<th>GOOSE</th>
<th>PRICE</th>
<th>CHOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>concern</td>
<td>mountain</td>
<td>raising/racing</td>
<td>homes</td>
<td>soon</td>
<td>cried (2x)</td>
<td>boy</td>
</tr>
<tr>
<td>third</td>
<td>down</td>
<td>safety</td>
<td>told</td>
<td>two</td>
<td>tried</td>
<td>time</td>
</tr>
<tr>
<td>heard</td>
<td>however</td>
<td>stayed</td>
<td>go</td>
<td>zoo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>out</td>
<td>louder</td>
<td>gave</td>
<td>same</td>
<td>escaped</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CLOTH</th>
<th>SQAURE</th>
<th>NORTH</th>
<th>FORCE</th>
<th>CURE</th>
<th>USE</th>
<th>happy</th>
</tr>
</thead>
<tbody>
<tr>
<td>long</td>
<td>air</td>
<td>short</td>
<td>forest</td>
<td>before</td>
<td>unfortunately</td>
<td>company</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>unfortunately</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>exactly</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BATH</th>
<th>START</th>
<th>THOUGHT</th>
<th>letter</th>
<th>NEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>afternoon</td>
<td>dark</td>
<td>thought</td>
<td>bother</td>
<td>near</td>
</tr>
<tr>
<td>after</td>
<td></td>
<td></td>
<td>later</td>
<td>fear</td>
</tr>
</tbody>
</table>

Table 3.4: Token distributions according to lexical sets from The Boy who Cried Wolf (items/tokens = 90 per speaker)
3.3.3 The Sociolinguistic Interview

‘The basic tool for recording conversation in sociolinguistic variation is referred to as the ‘sociolinguistic interview’. In fact, this is a misnomer; a sociolinguistic interview should be anything but an ‘interview’ (Tagliamonte 2006:37).

Unlike the other styles, casual conversations or ‘the sociolinguistic interview’ is less structured, highly flexible and open-ended, and traditionally still the commonest approach to collecting a corpus of naturalistic data (Milroy & Gordon 2003:57). Considering its prime goal – which is to extract the most honest sample from the speakers, my objective was to get the respondents talking for as long as possible, and enough for a sizeable amount of clean and useful data. Hence, my intent was essentially ploughed towards setting off discussions on particular topics each speaker would find fascinating, while being an eager listener or a keen learner. Due to the ostensibly loose structure of casual conversations, the literature is awash with recommendations on parameters including: the span of talk-time, questions to be asked, and interviewer’s poise.

Depending on the encounter, Labov recommends an hour to two of recording for a full range of demographic data from each speaker (1988:32). Since the key objective of the exercise is to ‘dodge’ the so-called observer’s paradox also known as the Hawthorne effect, it is reasoned that interviewees usually settle in to their most casual mode after about an hour of chatting, and become less conscious of borrowed voices. Despite the verity of such thinking, research has shown that many are, in fact, prone to varying structural trends in the course of long conversations. For example, in her study on style-shifting, Schilling-Estes (1998) notes that, depending on some tentative reasons, respondents do frequently fluctuate in favoured styles; thus making it difficult to ascertain the range of period through which the speaker can be assumed to have relaxed from being ‘observed’. This is much likely in the light of registers. Speakers may feel the need to use corresponding styles for varying topics as thought suitable, or may gallop through emotional highs and lows in the course of the interview; thus inconsistent. Barring such variability in speakers’ behaviour, an interview can be as short or as long, depending on the goal of study, as a conversation lasting between 20 to 30 minutes can provide a decent amount of tokens for phonological analysis (Milroy & Gordon 2003:58). Based on auditory judgement in the course interviews, I expected some correlation between speech styles and vowel realisations. But in spite of efforts towards collecting a range of corresponding styles, I noticed that some speakers, particularly of the younger group, watched
their pronunciations in the citation and reading passage styles more closely than in casual speech. But this didn’t appear to be a universal trend. There were those who sounded quite consistent in the way they spoke throughout the talk, while others actually coursed the stylistic undulations based on the topic.

Plausibly, Milroy and Gordon (2003:58) have noted that ‘while the goal may be to engage the subject in free conversation, the interview situation is very different from the spontaneous discussion that might arise among friends’. And that ‘the responsibility for keeping the conversation going rests with the interviewer who manages the discussion by asking questions’. Some of the important factors to be considered in designing such question grids include the community and age of the speaker (Tagliamonte 2006:38). In a module for speaking with young speakers credited to Labov (1984:32-34), the order follows through from basic questions on: demography, hobbies, witnessed fights, dating and marriage to danger of death experience(s), dreams or nightmares, religion, peer encounters, school experiences and language (Milroy & Gordon 2003:59).

It follows that while questions in the sequence of: the school in the neighbourhood/attended, its distance from the house, relationship with teachers at school/with parents at home, Jocks/Nerds/Goths/Thugs membership and so on, might just be fitting for an adolescent or a young eventful speaker; many older speakers would either become overly offended or shirk them all together. It is the same for religion and culture. Among Nigerians, ethno-religious sentiments are deeply wedged into people’s identity. My fieldwork was carried out among a highly religious population. Though educated, many of my adult respondents had strong opinions either about their native norms or religion. I interviewed a male polytechnic Admin Officer who kept code-switching between English and Classical Arabic throughout the interview. Telling such speaker I preferred his expressions in English might have been heard as not merely offensive, but also irreverent. In view of these complexities, keeping them unflustered required adapting a flexible structure – being aware of what to say or otherwise and how to say it, hence the need for a talk frame.

In sketching out a talk frame, I relied on Labovian adaptations in Poplack (1989), Poplack & Tagliamonte (1991), Tagliamonte (1997), Tagliamonte (1999, 2005). My golden rule was to second-guess the brand of conversation each speaker might enjoy, and privilege it through the peak of our talks. Before my fieldwork, I read severally through the modules as revised so as to localise them. I noticed that while certain questions only required modification for certain speakers, some were completely unusable. For instance, talks about
family meals (Module 9), though seemingly evoking – especially with speakers in Western climes, would fail among folks who have no such tradition of eating dinner/breakfast together. Also, while my female respondents rarely had trouble picking a favourite meal and giving detail of how to make it, the majority of the younger males either could not tell the meal they cherish most or simply gave a name and kept quiet. And on the thought that such questions about kitchen skills might offend my older male speakers, I excluded it for the group.

Due to the social ecology in which a lot of younger speakers were raised, talks about dating (Module 4) often stay clandestine, and this was possibly why majority of this group, especially females, were not predisposed to disclosing their personal adventures. Again, it would be rude to mention any such with an older speaker! On the other hand, the younger ladies were jolly talkers about their dream matrimony, career goals, and best friends or roommates at schools, their childhood and relationship with parents. Nigerians are generally informed on burning political matters around them. Issues bordering on the state and politicians roam freely, and every adult often has something to say. I had particularly interesting sessions during which speakers had seamless flows on politics till they got obviously emotional. An older Muslim female speaker recounted a nostalgic memoir of her teenage years in a polygamous family, and vigorously lamented what she explained as the danger of growing bias against polygamy in the wake of modernism.

As parts of the ethical standards, each speaker was well prepped up – basically to diffuse suspicions and unease. In the event of doubt or on request, speakers were also informed of their right to demand exclusion from the study or deletion of personal information from the recordings. This awareness, logically, turned out to be a little knotty. My ordeal was similar to Gordon’s, while interviewing a café owner during his fieldwork in Michigan. The speaker, on suspecting the purpose the answers would serve, declared, ‘You don’t care about this stuff; you just want to get me talking’ (Gordon 2001b, quoted in Milroy & Gordon 2003:62). Seen with recording gadgets, some of my respondents were spurred on to share their minds mainly on the thinking that ‘what’ they had to say; not merely ‘how’ they said them, would interest me. A few managed to realise my intent, but had concerns about their privacy. With the former, I took on the poise of an interviewer, a keen listener, and a fan of what they were sharing but spent quite some time explaining to the latter that I would, in the end, be minding only the structure of their samples. An older female speaker had called me over to her office on the following day of our interview to demand further clarification on what I planned doing with her ‘stories’. During her interview, she had grown deeply
emotional while telling of suffered betrayals from a family member, plunging into tears. At this point, I paused the recording, and asked afterwards if she would still like to discuss some cheerful topics instead – to which she agreed. Apparently, my re-invitation was towards convincing her I was not an undercover agent posing as researcher, and to double her confidence on the safety of her personal details. Younger speakers however had no such concerns. While the older male respondents enjoyed discussing politics, majority of the younger cohort were soccer fans. In fact, it became the norm to ask speakers in this group for comments on the most interesting games they felt their favourite teams ever played.

3.4 Recording

My field recording, no doubt, had its odds. Finding noiseless spots for relatively clean recordings and power for my device were parts of the hurdles. So, I went equipped with an H4n Zoom handheld recorder with an AUDIO TECHNICA AT8531 lavalier microphone. It is a dual input handheld state recorder that accepts either XLR or TRS combo connectors, with inbuilt capacity to limit, filter, modulate, delay or amplify models. All recordings were stored in WAV format to avoid acoustic compression as well as guarantee quality retention. The WAV files recorded by the H4n can be either 16- or 24-bit, with a sampling rates of 44.1, 48, or 96 kHz. My recordings were set at 24-bits with a sampling frequency of 41.1 KHz, which was later down-sampled to 16-bit in PRAAT (Boersma & Weenik 2013). Perhaps, the properness of H4n for my fieldwork is its compatibility with Audio-Technica AT831b, Cardoid condenser lavalier microphones – easily attachable to speaker’s chest, instead of head-worn microphones. This, in no small measure, minimised interviewee’s consciousness of being recorded or the so-called observer’s paradox (Di Paolo & Yaeger-Dror 2011, Hofmann 2015). Very fittingly, H4n has great battery life and recording span – such that in the absence of power supply (which was often the case); the device required just a pair of AA batteries that could actually last about 14 hours (if fully recharged). Recorded data were stored directly to a 32-gigs SD card, from which they were transferred to the PC.

3.4.1 Segmentation and Measurements

Parts of the procedures in the overall course of extracting tokens from the data involve segmentation and measurements. I begin this section by detailing the methods used and decisions I had to make. The essence of automatic segmentation through forced alignment is
integral to sociophonetics studies – especially those involving a large amount of corpora, and when more than just a few set of speakers’ vowels are under investigation. The process is specifically trained to fulfil the correlation between linguistic categories or levels (such as lexical items, phones and phonemes, and phrasal structures) and their acoustic incidents.

Depending on the goals of analysis, linguistic categories are reliably mapped unto the corresponding signal constituents, which are then sliced into segments and labelled accordingly. A manual alternative to this process obviously posed a measure of limitations, considering the scale of my data. First, it is subjective as the labelling of all phonemic units would rely on auditory evaluation and judgement. In addition, it would be unbearably time-consuming, especially with the segmentation of all tokens in the informal speech for all speakers. To mitigate these shortfalls, I turned to MAUS (Munich Automatic Segmentation System) for generating the Textgrids – from which tokens were extracted. The MAUS algorithm combines simple forced alignment based on Hidden Markov Modelling (HMM) with ancillary mesh of statistical enrichment to accommodate specific variants of languages (Strunk, et al 2014:3940). It is equipped to locate the most suitable correlation for speech signals or acoustic properties fed into it based on pre-trained parameters for the language(s) in question. Its predictive potential however depends on the quality of inputs (audio recordings and texts) which it basically interprets against all embedded signals and categories. MAUS, as currently implemented, supports about 14 different languages which include English and independent SAM-PA that automatically makes use of equivalent HMMs to yield best relevant approximations for the sounds in question.

To successfully feed my data into MAUS, certain conditions had to be met. First was text normalisation which required the transformation of all abbreviations, figures and dates, and the removal of special characters such as punctuations, and so on. Bearing this in mind, I adopted MAUS-compatibility rendering of the sound-files during transcription, as well as tab-delimited formatting before feeding it alongside the audio files on WebMAUS for segmentation. For each speaker and style, this process yielded separate Textgrids for the marked length of utterances. Since MAUS was not specifically designed to handle phoneme labels based on lexical classifications, i.e., Wells’ Lexical Sets (Wells 1982); a clever add-on which helped ‘clean up’ the tiers and further aligned the vowels with Wells’ paradigm was run on the Textgrids, thus yielding Textgrids that clearly matched up with the tiers and phonemes. The MAUS Textgrids MAUS, however, was by no means perfect, as some phonemic boundaries could be imprecisely read or wrongly labelled. As would be expected, it
remains far from possible for acoustic models to account for all the variation that could be present in the speech signal; but for studies as this one, a small degree of error in the phone alignments is acceptable, as long as the formant extraction procedure can ultimately obtain accurate results from the output (Evanini 2009:52). For further cleaning, I imported each sound file with its Textgrid into PRAAT for audio-visual inspection as well as adjustments and re-labelling prior to tokens extraction. The automatic coding system alternative was thus time-saving and helped immensely to avoid conceivable human errors typical of manual segmentation.

3.4.2 Formants Measurement

The question of the points at which tokens might be measured is crucial and often constitutes a major procedural concern in the analysis of vowel formants. After final adjustments and correction to phoneme boundaries on PRAAT, the next procedure was to determine the points at which to extract measurements for vowel tokens. Very imperative to discussions on formants’ measurements and their points of extraction are some different, but related accounts of vocalic spectrum known as *vowel inherent spectral change* (VISC). Generally, the arguments all take into account formant frequencies as central to the perception of vowel quality, but differ on what additional signals genuinely actually plays additional role in listeners’ perception.

While the onset + offset model privileges the properties of vowel transitions, the onset + slope hypothesis considers only the rate of change (i.e., the time or speed of the formant circle). The onset + direction hypothesis however relies on the broad patterns of formant movement (excluding the speed or time) as important (Nearey & Assmann 1986, Morrison & Nearey 2007). Though these paradigms (Figure 3.3) have so far only been assessed on simulated and natural tokens, they have wider implications for vowel measurements – in so far decision on the best model definitely affects the quality of tokens to be used for analysis. Based on a listening test of 408 trials each for 23 speakers (Morrison & Nearey 2007), the onset + offset hypothesis significantly outperforms other models in determining vowels’ spectral complexities. Their recommendation of multiple-point measurements is, thus, not surprising especially in the case of traditional diphthongs which can only be fully captured by measuring the nuclei (onset) and the glide (offset). All the same, the consideration of speed and formant movement in F1 & F2 space by the slope and direction hypotheses is equally
crucial to the overall detailing of the formant structure. For instance, the use of equidistant temporal locations and relative space for measuring vowels trajectories at multiple points (right from the onset to offset) with the aim of sampling every cue has featured prominently in major dialect studies (Hofmann 2011, Brato 2012, Baranowski 2013).

Figure 3.3: F1 & F2 formant features of natural and synthetic vowels. The comet plots labelled /e/, /i/ and /ɜ/ represent the mean formant onsets and offsets of the vowels as produced in isolated-word /bVp-/ context in Western Canadian English. Comet heads and tails mark the formant values at 25% and 75% of the vowels’ duration (from Morrison & Nearey 2007).

The single-point measurement approach, usually at midpoints or steady-states (points of inflection) was historically sufficient for vowel analysis until studies were carried out to justify the influence of extra phonological cues for vocalic identification (Morrison 2007, Di Paolo, Yaeger-Dror & Wassink 2001:90-94). With the single-point measurement approach, the decision as to where it is best to extract formant values is by and large arbitrary.

The golden rule of thumb, therefore, is to take the measurements right on the vowel’s steady state – which is the point at which formants have fully recovered from surrounding cotextual influence (Ladefoged 2003, Evanini 2009: 59, Di Paolo, Yaeger-Dror & Wassink 2001:91, Hofmann 2015: 167). The extent to which this is prudent would depend on the
research questions, as study with goals to explain lectal variations based on speakers’ *locus equation* in a particular language or dialect would rely on formant behaviours within transition at onsets (Delattre, et al 1955, Sussman 1991, Thomas 2011:102-3). Unlike the single-point measure at mid-point, the multiple-measurement approaches are fairly capable of covering the entire trajectal trail from the onset through the offset, including the steady-states and contours. Several measurements could be taken from a fixed default distance into the vowel based on certain metric rules. They could also be taken at regular intervals, i.e., in milliseconds (see below for further discussion on the regular interval approach). The proportional distance approach however takes measurement for every token in the corpus at similar points (mostly three points) along the trajectal trail, hence allowing for overall metric similarity for the vowels (Di Paolo, Yaeger-Dror & Wassink, 2011:91-3). These methods, thus, provide very useful access to the analysis of diphthongal vowels as well, and can sufficiently capture the internal dynamics of monophthongal behaviours (especially nasalised vowels) right from onset to offset.

To salvage time and resources, as well as lessen chances of subjective Hertz assignment for vowel formants, measurements were automatically taken at multiple-points with PRAAT script – from which I eventually relied on the midpoint or steady-state measurement for monophthongs; and on the onset/glide measurements for diphthongs (Evanini 2009: 58). To monkey around this, I determined a percentage (50% at midpoint) for monophthongs and at (20% & 80%) for diphthongs for formant Hertz measurements. The multi-points parameterisation is however, not without flaws, as evidences abound against its descriptive adequacy of all relevant spectral properties (Morrison and Nearey 2007, Morrison & Assmann 2012:41). However, in order for inter-trajectory comparison across speakers, some studies in lectal variations actually recommend as much as ten-point parameterisations using either a fixed range of milliseconds or percentages throughout the phone duration (Di Paolo, Yaeger-Dror & Wassink 2011). Often, this is done so as to show the complexity of formant contours for a sole point of inflection (Peterson & Barney 1952:177), and on the assumption that listeners often take into account the entire gamut of formant complexities for perceptive cues.

Despite lack of consensus as to the exact points of measurements and vowel-specific modifications, a two-point measurement approach – at 40ms into the vowel and 40ms before the closure was used in Andruski & Nearey (1992); and at 20% and 70% of the spectral stretch in Hillenbrand et al 2001); 20% and 80% of the vowel spectrum (Hillenbrand and
Nearey 1999); and at 35 ms from the onset, 50 ms (at midpoint) and 35 ms before the offset (Kendall and Thomas 2012). As noted in Hofmann (2015:168-169), none of these approaches is without practical shortcomings. For instance, vowel measurements at fixed distances through milliseconds would be unsuitable for a sociophonetic data since the phones would most likely have unequal duration – a common scenario with corpus containing a mix of tokens from formal and informal speech. The approach basically yields a number of readings that are neither too close nor apart among different durations, thus making internal comparison of different numbers and time-points very difficult (Thomas 2011: 153). Since the specified intervals approach disregard duration across tokens of unequal length, shorter vowels would likely take the hit while formants from longer vowels may differ meaninglessly from typical values. Hofmann further explains why this method may be unfitting for measuring a sociophonic corpus:

Assuming that the shortest vowel is 70 ms in duration, the earliest measurement point may be at 20 ms after the transitional phase of the formants from the preceding consonant, and the latest measurement may be 20 ms before the transitional phase of the formants to the following consonant. This would leave a measurable duration for that vowel of 30 ms, resulting in, for instance, three points of measurements at every 10 ms or less. The number of intervals is, however, much larger for vowels 200 ms in duration, and yet every point of measurement after the first three is unusable in a comparison of formant trajectories (Hofmann 2015: 184).

In view this deficit, the interval approach hence falls short of the criterion. For monophthongs, selecting a measurement point at the vocalic midpoint based on percentage is relatively common and poses little or no procedural complications (see Steinlen 2002, Pierrehumbert et al 2004, Chen et al 2009, Evanini 2009, Thomas 2011, Brato 2012 and Hofmann 2015). First, the use of percentages instead of milliseconds absolves all measurements of durational compromise. It also mitigates, to a large extent, the undershoot effects from both ends of transitions. Non-monophthongal vowels, however, often consist of more than one steady state, thereby showing dynamic formant contours, i.e., in the nucleus and the glide. Because of this, a diphthong may show inter-regional transitions between the onset and the nucleus, between the nucleus and the glide, and between the glide and the coda or the following environment (Thomas 2011:172). As a result of consonantal contexts or lectal variations, some traditionally diphthongal vowels may suffer truncation and realised as monophthongs by certain speakers, thereby showing just a steady state. An instance of this,
based on the auditory evaluation of my data is the non-formal realisation of FACE & GOAT tokens by speakers (Figure 3.4) – in which the glides are inherently absent.

![Figure 3.4: Textgrid showing monophthongal realisation or absence of glide in GOAT (in informal speech (MUO1)) & FACE (in reading passage (FUY2))](image)

To quantify steady states in a diphthong, measurements could be taken at close time-points across the whole vowel trajectory. The differences between time-points within a steady state, if it is present, ought to be very minimal while a larger difference reflects the presence of complex contours or absence of steady state. Hertz values obtained from these may then be compared across both social and linguistic categories for lectal variations or linguistic patterns (Thomas 2011:174). However striking this approach may seem, its potential ranks low, especially with a large amount of data. For typically diphthongal vowels, a two-point measurement including the nuclei and the glides has proved very sufficient in showing the dual trajectories (except in cases the vowels’ diphthongal status is under investigation) is usually enough (Thomas 2000, Labov, Ash and Boberg 2006, Evanini 2009: 60).

### 3.4.3 Data Cleaning

Initially, a total of 90668 vowel tokens were imported from the segmentation onto a spreadsheet with an add-on PRAAT script – Track Vowel V2 suite. 78685 were from the casual speech, thus contributing about 85% of the data. 7678 and 4305 were drawn from the reading passage and wordlist data. Since my commitment, primarily, is to a typical representation of vocalic structure, I excluded all vowels in unstressed and semi-vowel
contexts, and those in function or tapered words – to start with, trimming the tokens to 26987. Figure 3.5 is a snapshot of this procedure. Vowels flanked in either following or preceding approximant contexts, i.e., /w/, /j/ (except in the case of fronted USE), /r/ were separated from the big table and preserved in a subset for further examination of conditioning effects.

Prior to the extraction of data from PRAAT Textgrids, formants blurred by intruding noises or speakers’ involuntary actions (such as emotional outbursts during personal narrations or accidental interference with the microphone) during recordings were segmented and coded for exclusion. Left with likely tokens of interest, each Lexical Set, especially monophthongs were grouped and the F1&F2 values were set in the order that easily reveals the atypical formant values. Odd formants could haze up patterned variations among speakers and result wrong statistical outcomes, hence the need to meticulously exclude them from the data, especially when vowel shifting is not the core of the investigation (Thomas 2011:158).

Figure 3.5: A spectra cut of a speaker at the beginning paragraph of the The Boy Who Cried Wolf. The first 4 tokens in unstressed and approximants were coded for exclusion while the next 2 tokens (CURE & DRESS) were retained. The last (lettER) only occurs in unstressed syllables – hence retained.

Drawing on the only known acoustic approximation for an Ebira L1 speaker (Ladefoged & Maddieson 1996:304), and the typical formants average for English vowels (Ladefoged 2003: 114), I estimated a sealing for the monophthongs in F1&F2 – against which the tokens were initially examined. For high vowels, for instance, formant values for an adult Ebira male speaker range between 280 Hz – 350 Hz in F1, and 2000 – 2400 Hz in F2; while the normal range of
high back vowels peaks between 250 Hz – 300 Hz in F1 and 600 Hz – 700 Hz (Ladefoged & Maddieson 1996:305). Apart from tokens obviously conditioned by phonological environment or other linguistic variables, those found to be outrageously weird at mid-point in F1 & F2 were inspected against corresponding bandwidth measurements. As a rule of thumb, smaller bandwidths are indicative of accurate readings by the LPC algorithm, except for nasalised sounds. Normally, a sharp peak at a low frequency would have a low mean frequency, but a shallow peak at an equally low frequency would have similar mean value but with corresponding large bandwidth (Reetz & Jongmann 2009:159). Thus, for non-nasalised tokens, bandwidths above 400Hz were re-inspected for possible conditioning effects of linguistic factors, and in instances this was not the case, I discarded the tokens completely from the data for a relatively purer table.

For a better scan through the values of tokens in F1&F2, the quantile range of dispersions was used. I filtered each speaker and vowel class into separate excel sheets and had their quantiles taken in R (RStudio Team 2015) at 5%, 10%, 50%, 90% and 95%, thus yielding discrete distributions at these intervals. This proved to be much more effective – in that, dispersions from central tendencies became more detectable, and could easily be checked against accompanying variables (i.e., linguistic environments) and other acoustic information – to determine their weight of accuracy. Rather than knife off tokens outside the closest 5% and 95% quantiles (Turkey 1977, Thomas 2011:159), the total counts of the tokens in question and their phonetic contexts were re-checked. Following these procedures, I based decisions on whether to exclude such tokens on comparison of the occurrence of same vowel in similar phonological context for the same speaker; and in cases where values could not be explained by characteristic factors, they were excluded.

However, since quantile estimates of the smallest and largest ranges for each phoneme and speaker could only help trim off extreme formant values, further visual inspections was necessary to detect the other outliers. Based on corresponding formant values, therefore, I re-checked the spreadsheet for data points visibly detached from the mean density on scatterplots.
Figure 3.6: F1 & F2_normalised plot for all monophthongs prior to final cleaning and removal of tokens of less than 70ms and above 220 ms from the dataset (n=14905)
This procedure was repeated individually for all speakers across the vowel classes, and tokens perceived to be meaninglessly away from the central density were removed. After these cleaning procedures, my tokens were sizeably down from 26987 to 22735. An F1/F2 scatterplot of each monophthong revealed considerable dispersion towards the central region of the vowel envelope. Conceivably, those tokens which angle extremely from their peripheral domains might have suffered durational reduction thus rendered schwa-like, owing to prosodic or undershoot effects (Thomas 2011:172).

Consequently, tokens within this region of the plot must have had their steady states truncated or completely lost to flanking phonological contexts. A likely knock-on effect of these tokens, if left in the dataset, would include: wrong impressionistic observation and Type 1 & II errors in further statistical assessments. In cases of extreme discrepancies in phone duration – caused by prosodic factors, it is usually advisable to normalise for rate of speech, especially before examining contrastive or contextual vowel length (Thomas 2011:144). Another would be to exclude tokens with comparatively low and high milliseconds, since their steady states are not expected to be representative anyway. Thus, following Wassink (2006), I normalised duration for the tokens, and also excluded tokens with duration below 70 and above 220 milliseconds. This further reduced my total number of tokens to 18740.

The foregoing has mainly shown some of the procedural considerations involved in the design. Given the goal of clearly outlining a variationist procedure that can be replicated for future studies, the the constitutive aspects of speakers’ grouping into cohorts as well the layers of sociolinguistic interview were extensively explained. Critical to the categorisation of speakers, especially, were the factors of age, gender and education. Though only speakers who had had some tertiary education training were included, the subsets of speakers trained in the polytechnic and the university was created – so as to assess if such differentiation would indicate an effect. Also discussed in this chapter is the choice of elicitation resources – which include the key determinants in the construction of questions for casual conversations. These sections have, very generally, described the manner in which the data was collected. The following chapter thus discusses the some sociolinguistic theories which I consider relevant to the evolution of NigE varieties. In what follows, I expand on the determinants in choosing the normalisation as well as the statistical tool prior to the analysis in Chapter 5.
4 Theoretical Frameworks

4.1 Introduction

As discussed in Section 2.1 & 2.2, the emergence of NigE, particularly in Ebiraland is a consequence of diverse antecedents – some of which delimit broadly within linguistic and psycho-social factors. In this section, I reconstruct a range of theoretical scaffolds from the variationist frameworks, on which the nuances of the variety might be explained. Since the grid of this study bestrides a cluster of factors, I adopt a rather eclectic or multi-conceptual approach that would include its overall descriptive goals. The choice of each design therefore lies in its resourcefulness or aptness to the analysis of Ebira English as an L2 variety.

Considering various theoretical fronts, it is fair to note that a few of the earliest works on language contact and the L2 experiences differ fairly from some current preoccupations in mainstream sociolinguistics. Until recently, explanatory tools and theoretical standpoints were largely derived from language acquisition models, focused mainly on individual speakers rather than a group of language users (Selinker 1972, Richards 1972, Schumann 1974, Sankoff 2001, Weinreich 1951, Ferguson & Gumperz 1960, Adamson and Regan 1991, Tarone 1988, Preston 1989, Major 1998). This basically syncs with the variationist traditions, especially in investigating variation among individuals or speech communities. It is thus reasonable to emphasise that the range of our current awareness in sociolinguistic variations also broadly synergises with concepts in the traditional L2 linguistics. These interactions have not merely built a strong theoretical lattice for the analyses of language use in contact situations, but have also underpinned the variationist enterprise. In her synopsis of possible linguistic outcomes from contact situations, Sankoff stresses the exceeding influence of the phonological substratum and its exposure to evolutionary pressures – the consequence of which does remain as long as the new varieties (L2s) make it through generations of speakers (Sankoff 2001).

Further designs that explain the gradients of L2 realities have consolidated on these concepts, including supplementary proposals for better understanding and prediction of consequences in emergent bilingualism. Prominent of these are Schneider’s (2007) Dynamic
Model for postcolonial Englishes, and Mufwene’s (2001) notion of language ecology, which both underscore the complexities of linguistic trajectories in contact situations. The descriptive adequacy of these frameworks essentially lies in their inclusive potential of both internal and external factors in the overall dialogue between the host and the new structures. For instance, coupled with incipient factors such as trade and missionary activities, English in Nigeria was originally a colonial exigency – which is why a theory for its descriptive project should conceivably build on this premise. Drawing on Thomason (2001) and Winford (2003), Schneider has indeed outlined some relevant aspects generally involved in the construction of Post-Colonial Englishes (hence PCEs). They include: the intensity of social factors activated and sustained through the early stages of linguistic perpetuation in the colonies; the correlative links between the linguistic structure of the new variety and the social conditions of its speakers; the role of code-switching and alternation; reflexive awareness; the speakers’ primary languages; and individual idiosyncrasies (Schneider 2007).

4.1.1 Mufwene’s Feature Pool

Further views on contact have similarly been expounded in Mufwene’s writings. Adapting ontological parallels in biological concepts of evolution and ecology, he conceives variation as the perennial consequences of natural selections from competing alternatives being generated in the idiolects (and dialects) of individuals and communities (Mufwene 2003:146). He finds that selections which are mainly consequent on contacts at different levels are most responsible for the emergence of new varieties, and that below the macro-level of contacts between languages or dialects is the more incipient contact at the micro-level, i.e. between idiolects. Thus, events of language contact are as ever-present as the casual use of language within or between linguistic communities – as much as it equally follows that the course of linguistic change or variation begins with individuals who constantly donate their idiosyncratic features into a linguistic pool, creating a whirlpool for the emergence of new varieties. Mufwene approximates such vortex to a ‘feature pool’ (analogous to ‘gene pool’ in biology); and the process of appropriation of linguistic items/segments from the pool by individual speakers as ‘transmission’:
According to the feature-pool approach, partners in the reproduction of a species, which I take a language to be, make contributions to the structural makeup of new members (idiolects in case of language), which share features on the family resemblance model. These new members select their features (typically with some modification) from the same pool, although the recombinations are never the same, and those which wind up as dominant are not always the same. In those cases where the contributors have more or less the same (kinds of) features, it is pointless to try to identify a single (dominant) donor (Mufwene 2002:47).

A major challenge with these analogies lies largely in conceptualisation. For instance, while languages are either acquired or learned, genes are taken from fore-parents. Mufwene is aware of this ontological difficulty, but argues that his explanation of language on population genetics assumes the former as species rather than organisms. Unlike the wholesale passage of genotypes to the passive recipients, the transmission of linguistic items from the feature pool is achieved with varying adjustments, such that ‘systems are not passed on intact from speaker to speaker’ (Mufwene 2003:147). In the traditional cycle of reproduction, a gene passed on from a ‘donor’ to an offspring is perfectly replicated in the offspring. This is hardly ever the case for any human language, nor acquired this way. In the process of acquisition, though speakers do apparently grow with the language, they are far from being passive receivers of its supersystems (Mufwene 2003:208).

Language acquisition or learning involves some form of recreation or reconstruction, hence imperfectly replicated (Hagège 1994, Lass 1997). So it is with speakers’ behaviour within a speech community. A speaker may take a liking to a particular idiolect or dialect or even begin to consciously accommodate its styles as a mark of identity/association. Processes of such can initiate a radical change in the individual, a group of speakers (with common social features), or a wholesale change at the communal level. In fact, the crux of most studies in sociophonetics is concerned with investigating these fallouts. Often, studies are designed to investigate intra- or inter-group variations within populations – a logic that assumes the trend of accommodation among speakers with shared social features and plausible variations among different groups. Mufwene refers to such determinants, both linguistic and social, as the language’s macroecology – which, originally, is a biological construct that accounts for a range of influential factors (internal and external) responsible for species’ evolution (Mufwene 2003:153). Evolution is therefore subject to sociolinguistic constraints. A group of individuals may either feel socially conditioned or condition themselves to speak in a certain way, or favour the use of certain codes that define them from other speakers.
With regard to language, two evolutionary tendencies have been identified. One entails either partly or a radical progression, a change towards complexity or structural novelty, while the other is usually brought about by natural selection, i.e., occurring without an attendant progression of any kind from impure realisations to more satisfactory ones, or necessarily from an easier to a more intricate system or vice-versa (Nichols 1994). This other type of evolution, according to Nichols, is Darwinian – a *speciation* process of some sort. By speciation, Nichols refers to a subtype of Darwinian evolutionary process by which reproductively isolated biological populations evolve to become distinct species. However, she is ostensibly persuaded of no such similitude in linguistic, consequently stating that though ‘linguistics has no analog to the biological notion of species, it’s safe to say…that languages are related to each other as individuals or kin groups of a biological species are’ (Nichols 1994:276).

On the analogous possibilities for speciation in linguistics, Mufwene’s stance is rather affirmative. He maintains that linguistics has parallels to certain components of speciation in biological species (Mufwene 2002:46, 2003:146). For instance, he weighs the effects of inter-idiolctal interactions within or between communities, and the role of natural selection in the choices speakers make. He however dismisses the notion of progressive evolvement in the process of linguistic evolution, and affirms that ‘evolution has neither a certain purpose nor pre-defined goals. Linguistic change, therefore, is unintended, a consequence of imperfect replication in the interactions of individual speakers as they adjust their communicative strategies to another or fresh needs. Resulting from such trends could be: a communal recycling of structural possibilities, generalisations due to structural congruence, increased irregularities, introduction or obliteration of useful distinctions (Croft 2000, Mufwene 2001:12, 2002:47). The very notion of ecology as the cloud of socio-linguistic determinants underscores the vagueness of linguistic evolution – given the natural tendency of language features to change towards either structural complexity or simplicity, or neither of both (Mufwene 2003:147). Evolution thus involves the different stages of ‘restructuring processes’ through the history of a variety; that is, the modifications of the source structures (the phonemiser) or between-speaker accommodations within a system. By *restructuring*, he thus foregrounds the roles of ecological triggers in variation, and their prospects for linguistic and social factors (see Section 5.4).
4.1.2 Linguistic Species and Accommodation Theory

Perhaps, the most interesting core of Mufwene’s equivalence of linguistic processes in biology is the idea of a language as species than an organism. As an organism, reference is made to individual styles, i.e., an idiolect. But this is not sufficient, as any variety actually consists of different idiolects. Just as more than one organism doesn’t make a species, a language is an extrapolation from idiolects governed by similar underlying systems (Mufwene 2003:149). A logical parallel is found in the Chomskyan delineation between the internalised languages *I-languages* and the externalised languages *E-languages* (Chomsky 1986).

Basically, while the former makes up the individual speaker’s language, i.e., a lattice of linguistic principles in the mind, the other belongs collectively to the speech community, thus implying that every fluent individual in a language community has an I-Language, and as such can potentially produce an infinite E-Language. The E-Language is thus epiphenomenal since it essentially depends on the I-Language. He classifies the linguistic knowledge that is in the mind of the speaker as I-Language whilst the observable linguistic output (sentences, songs, texts etc.) as E-Language. Mufwene then approximates these as useful parallels for idiolects and dialects, in the sense that an idiolect is to a variety what an individual is to a species in population genetics; and that – since the region of dialect or language contact is the speaker’s mind, ‘the difference between idiolect and language or dialect contact is more quantitative than qualitative (Mufwene 2003:150).

Such differentiation is therefore important for our understanding of accent variation, especially in the sense of lectal dissimilarities among speakers. It explains, essentially, the commonness of by-speaker idiosyncrasies in the light of random factors that are not universal to certain cohorts or groups. Mufwene attributes the major concerns of variationist sociolinguistics to the customary notion of language as an *organism* – a view that continues to privilege comparison between dialects over idiolects in speech communities (2003:148).

The core of many studies in sociolinguistics often builds on the correlation between marked social indices and linguistic structures. At times, such factors are assumed uniformly shared within the group, with the exception of minor discrepancies usually dismissed as socially insignificant or statistically weak, therefore blurring our understanding of in-group variations. For instance, a correlative evaluation of sound patterns within a speech community may depend on factors such as: age, gender, education and other sundry predictors for the speaker groups, but not the individual speakers; and results from such measurements could
lead to general assumption about the speakers’ linguistic behaviour. Mufwene considers this an oversimplification and defeatist to the basic aim of variationist studies, hence regarding the idiolectal variants as more structurally consistent than communal behaviours:

With all the variation that is typical of communal languages, it may turn out that there is more systematicity in idiolects than in communal languages. Systems are needed by individuals, and in idiolects, for consistency in individual behaviours. It is all right when they translate into the communal system, but it is not necessary that they do (Mufwene 2003:148).

Systematicity therefore assumes greater similarities among idiolects of the same dialect than dialects of the same language. That in a population sample, subjects who share similar social characteristics would show lesser variation than those in separate categories. And that certain classes of linguistic items, e.g., front, back, low vowels would behave similarly in certain contextual environments than others. A striking nexus exists between structural fidelity at the idiolectal level and accommodation within the speech community. Since linguistic change tends to begin at the speaker’s level – perhaps a consequence of innovative donations by individuals to the pool of competing features, it is expected that the nature and trail of change depend largely on the surviving features (Giles & Smith 1979, Milroy 1992, Mufwene 2003).

Such change, however, is hardly ever rigid, as it could be a regular (internal) kind of change or externally motivated by contact with another tongue. Either of these is an outcome of underlying struggles between the competing systems within the speech community. Since idiolects and dialects are mutually exposed to gradual and unconscious accommodation through co-existence, the bias of choice or the ease of adjustment in the selection process determines which of the idiolect(s) survives in due course. This explains, in part, some reported cases of levelling and heavy linguistic borrowings in some parts of the native English speaking communities and former colonial territories (Hinskens 1998, Trudgill et al 2000). Through accommodation, some features gain selective advantage over other competitors which are selected out. In some cases, a network begins using a feature which is more typical of a different network even when most of the members of the two networks do not interact with each other (Mufwene 2003:151).

There is thus a conceptual dialogue between Mufwene’s linguistic equivalences in biology and Schneider’s Dynamic model (Figure 4.1). Both underscore the speakers’ selective potential from the pool of linguistic facilities available to them; the overall linguistic and social conditions that constitute their ecologies, and the possibility of a competing variant
becoming most pervasive. Depending on the nature of struggle for survival, the winning features might entail both: diffusion – vertical and horizontal transmission of items into the new variety; and selection – innovations adopted from an indigenous or the host language (Thomason 2001, Schneider 2000b). A process of this sort would keep the structure of the new variety (PCEs) perceptibly unstable at the early phases of evolution, while the structuring forces of the early vernaculars also remains profound (Mufwene 1996b, 2001b, Ajani 2007). Amid these is the prospect of lectal variation through replications, depending on speakers’ ability to reproduce the pool of features available to them, and the tendency to recreate and appropriate the elements of the new variety.

![Figure 4.1: A model based on Mufwene’s notion of language ecology and evolution](image-url)

### 4.1.3 Sociolinguistic Identity and Accommodation

It is often difficult to speculate narrowly on the sociolinguistic chemistry that results from languages in a contact situation. Outcomes such as: diglossia, language shift, attrition, code-switching, creolisation, pidginisation, birth of a new language or even linguicide (death of an existing language) are all thinkable (Ajani 2007). The contact of English with Nigerian languages, no doubt, has yielded more than one of these consequences. Beside desperate efforts on the part of the locals to preserve their linguistic heritage has been a competing
yearning to excel in the approximation of the English accents, and as a result, have held faithfully to their linguistic identity and at the same time accommodating the L2 features for sundry needs. The implications of sociolinguistic identity and accommodations, in many ways, have explanation for the social complexities of NigE systems (see also LePage and Tabouret-Keller 1985, Wodak et al 1999, Jenkins 2000, Schneider 2000b, Josiah 2011). Both explain the means (social and linguistic) speakers adopt to shift and conform to new systems and the natural compromise that surface from shifting between varying spheres of contexts, either within smaller or larger groups of collectives. This is true of NigE accents which comprise noticeably emergent, i.e., middle-of-the-road systems (Trudgill 1986, Brato 2012) for communication purposes among diverse ethno-linguistic formations in the country, and with the rest of the English speaking world.

In a proposal by Schneider (2007), he reasons that all accommodatory gestures often require some forms of ‘symbolic expression’, but also notes that those who find it difficult to demonstrate certain features marked with a particular social group might resort to easier ‘symbolic forms’ for their communicative needs. As such, group formation does not only possess social tagging, but also self-constitutive to regulate and determine its in-group membership. Thus, identity construction through linguistic means is constantly in a flux, subject to modification by ecological factors, i.e., the social tensions among the speakers. In a multilingual setting where hybrid representations are possible for instance, identities could overlap and conflict with each other, depending on the speakers’ social consciousness.

For an individual as well as a community, defining one’s identity implies a need to decide on who one is and, more importantly, wishes to be: a line is drawn between ‘us’ (those who share essential parts of a common history and value orientation, those who wish to socialise and be associated with) and ‘others’ (who are just perceived as different, and don’t share these qualities). Identity definitions entail both individual identification and social classification... Speakers who wish to signal a social bond between themselves will minimise any existing linguistic differences as a direct reflection of social proximity: they tend to pick forms used by the communication partner to increase the set of shared features and avoid forms which they realise are not used by their partner and might thus function as linguistic separators (Jenkins 1996:164-71).

In the course of transplanting an L2 into fresh territories, mutual alliance between the two systems is actively open to ‘negotiation’ (Thomason 2001, Schneider 2007), and structural compromise which allows for similarities and dissimilarities in the coding of budding varieties, depending on social and linguistic variables. These interactions apply in various
ways to the concepts of Schneider’s PCEs, and have theoretical links to his Dynamic Model. Basically, it assumes a shared underlying process in the evolution of New Englishes, and suggests a conceptual scaffold for the similarities between the clusters of New Englishes.

4.1.4 Dynamic Model

The Dynamic Model derives essentially from cognate theories of language contact including: bi- and multilingualism, L2 acquisition and L1 transfer, sociolinguistic variation, linguistic identity, and language evolution (Schneider 2014:10). Schneider believes that the emergence of English in new territories, with few exceptions, is characterised with similar patterns – that are mainly subject to sociolinguistic and other contact conditions, and that the historical mould in which Postcolonial Englishes have evolved are similar in trends and timelines (Schneider 2003). Depending on the manner of contact, the linguistic notion of ‘us’ and ‘them’ among the Settlers speech community (hence STL) and the Indigenous speech community (hence IDG) does undergo some form of renegotiation that could ultimately usher in a new variety (Schneider 2007:29).

Through its imperial history, English has transitioned from being a colonial language – to an elitist trademark – to the language of modernity, and a unifying emblem in linguistically heterogeneous countries like Nigeria, Singapore, Ghana, etc. In a bid to account for these developments, Schneider, in PCE, suggests a descriptive model (the Dynamic Model) that correlates with five evolutionary phases to new Englishes; namely: (a) foundation (b) exonormative stabilisation (c) nativisation (d) endornormative stabilisation, and (e) differentiation. And other factors responsible for inter-phasal progression – such as: historical events or language policies, the contact setting or ecology of contact, language use and attitudes, and the cumulative outcomes on these have on the emergent structures. Before proceeding on further discussion of the phases in the model, particularly as they relate to the evolution of NigE, a caveat from Schneider is in order:
Most importantly, I fully agree with Thomason that all generalizations relating to language contact are idealizations, like all models, abstracting essential observations from a messy reality but unavoidably ‘leaking’ in some respects: no typology in this area is exhaustive, and all possible generalizations may have to face a counterexample somewhere. Nevertheless, generalizations are possible, though only probabilistic rather than absolute ones… Nevertheless, it is important to emphasize that even if in specific circumstances some details may have developed somewhat differently and there may be apparent counterexamples to some of the trends worked out below, on the whole the process is real, and it is robust (Schneider 2003:241).

The foundation phase marks the virgin arrival of English on the colonies through the British Settlers on diplomatic assignment or as missionaries or traders – from the mid 16th century through the second half of the 19th century. At this phase, the speakers in the new territory are understandably few, as they include only the new Settlers and the earliest batch of the L2 speakers (Schneider 2007:200). In the case of Nigeria(ns), some major sociolinguistic events that trailed this period are discussed broadly in Section 2.1. I explained that the complexity of interactions inspired by the activities of the ‘visitors’ and the ‘locals’ at this stage must have occasioned the appearance of what is now popular as Nigerian Pidgin – a localised mesh of English and the indigenous languages. According to Schneider, some internal levelling of the STL strand ‘koinéisation’ could also occur at this stage. As earlier mentioned, the British expatriates and the Christian missionaries were first to speak English among Nigerians but did not share a homogeneous accent themselves (Awonusi 1986). Thus, to communicate successfully within their inner circle, and with the locals, some concessions might have been made at the most nascent periods of contact with the colony, i.e. by conceding their regional accents to those varieties with wider potential. By the same token, the locals who engaged in trading activities with the English also commenced a long process of linguistic concession at this stage.

During exonormative stabilisation (the second phase), the influence English spreads, becoming recognised as the language of government, schooling, legal system, etc., at least in some regions and class of society. As the number of new speakers among the IDG increases, imperfect replications of linguistic features would surface, paving way for a distinct variety. With regard to Nigeria, this stage lasted about forty years, from 1890 – late 1940s (Schneider 2007:201). As noted in Gut (2004:815), the majority of British residents in Nigeria at this period were the middle – and upper class RP speakers who initiated an orientation towards the Received Pronunciation (RP) or the Queen’s English among Nigerians. English was extremely class-defining, and those who could speak it were venerated (see Section 2.1). While English expanded in its formal functions, pidgin provided some form of ‘back-up’ for
casual and informal interactions among the largely unschooled population. This period also marked the earliest inflection of regional accents in NigE, yielding the nascent indicators of Yoruba, Igbo and Hausa English accents (Awonusi 1986).

The third phase is most vibrant and critical for phonological and lexico-syntactic innovations. It is mainly characterised with conflicts and resolutions, at the end of which the STL and IDG gradually come to terms with effects of contact situation. Crucial interaction between the old with the new means of communicating or talking is forged, giving way to fresh variants. The STL progressively succumbs to structural modifications – losing its original identity to the new speakers. Politically, this coincides with the period of independence from the colonial power. In 1960, Britain left the shores of Nigeria; leaving English behind – at the mercy of Nigerian languages, and as an imperial legacy which has continued to remain dominant and stronger in influence (Section 2.2). By this time, the indigenous variant develops some reflexive ‘idiosyncracies (through substrate effects, interlanguage usage, and the like), and in the ongoing mutual, if asymmetric, negotiation and accommodation process some of these will slowly be adopted by certain STL-strand users as an expression of their identification with their current country of residence, their future rather than their past, gradually supplanting their loyalty to the country of origin’ (Schneider 2003:248). With regard to NigE therefore, Schneider (2007) reports that though some evidence of nativisation is already pervasive, the variety is yet to fully attain the endonormative stabilisation. He however concurs that ‘the intensity of the nativisation of English in Nigeria is probably best illustrated by the reality of the literal sense of this word’, noting that there are Nigerians who actually acquired both English and Pidgin as L1.

At the next phase, the former colonies disengage from the imperial hold of Britain, and begin the journey to institutional and cultural independence. Schneider believes that though the transition to this stage might occur unnoticed, it is often anchored by mostly political events. He notes that while dialect birth is yet to become noticeable, the new variety becomes predominant with a unique singularity, and that in due course, the distinctive systems of the new variety undergo intrinsic formalisations and get wider recognition in the linguistic market. Gradually, the new variety receives indigenous and international endorsement, and begins to enjoy a growing sense of ownership among its new speakers. As it settles in and stabilises with the host languages, it also takes on a new label – switching from ‘English in X’ to ‘X English’.
The last phase – differentiation, is often ‘a consequence of eternal stability and internal cohesiveness’ (Schneider 2014:12). At this stage, English is expected to have surpassed a national homogeneity unto internal splits along social, ethnic, gender and regional divides, and the emergence of intra-variations thus becomes a plausible index of identity and group membership. Differentiation in the purest sense allows for style correlation with socially and ethnically defined speech communities coexisting under a common roof, andlectal variations amid a set of shared linguistic norms. He summarises the core markers of this phase:

Once a solid national basis has stabilized, one’s global, external position is safe and stable, as it were, and this allows for more internal diversification. The focus of an individual’s identity construction narrows down, from the national to the immediate community scale. The citizens of a young nation no longer see themselves primarily as a single entity against the former colonial power, but rather as a composite of subgroups, each marked by an identity of its own, determined by sociolinguistic parameters such as age, gender, ethnicity, regional background, social status, and so on. At this stage an individual’s contacts are strongly determined by one’s social networks, within which the density of communicative interactions is highest (Schneider 2003:243 & 253).

4.1.5 NigE in Schneider’s PCE

By the illustration in PCE model, Schneider assumes NigE as already on the threshold of fourth phase (Schneider 2007:210), but avers that, despite the progress made, ‘it would be exaggerated to claim that national unity and stability have been achieved’. This consideration is rather more consistent with the political rather than linguistic realities – a major point that exposes the paradigm to ontological deficits. Subsequent assessments of the model have contended its implicit binarisation of linguistic sequence, subjectivity and the clinical taxonomy of more or less dissimilar phenomena (see Mesthrie & Bhatt 2008:34-5, Ugorji 2015:32-44).

For the NigE varieties, the strongest reaction has been Ugorji’s (2015). But as I will point out shortly, his evaluation is likewise open for a review on procedural grounds. The placement of NigE on the brink of endornormative stabilisation in the model presupposes the absence of internal variations or dialect birth within the system. Schneider believes that ‘English in Nigeria has progressed deeply into Phase 3, and has nativised strongly, and is still gaining ground at rapid pace’, and even hints that the strand might already be on its way to
Phase 4 (Schneider 2007:212). For a variety at this stage, the model requires among other features, some form of structural convergence – to the level that it is hence regarded as X English (Nigerian English) as against ‘English in X (English in Nigeria). Ugorji however faults the model, noting that the variety has progressed steadily since the 90s, and argues that homogenisation and lectal differentiation co-exist in the Nigerian experience (Schmied 1991a, Jowitt 1991, Simo Bobda 2000, Udofot 2003, Gut 2004, 2007, Awonusi 2004, Jolayemi 2006, Josiah et al 2012, Ugorji 2015:36).

For differentiation – which is the last stage, the model suggests a dynamic level of internal variation and clear evidence of sociolinguistic markers within the country. Ugorji’s reaction to this resounds reasonably with the assumptions of this study: that the NigE variety was born with lectal differentiation right from contact (see Section 2.1 & 2.2) mainly discernible by layers of social factors among speakers:

... the commencement of lectal differentiation in the Nigerian experience can be specified... In general, it may be shown to have commenced with earliest contact inceptions. In particular, the Nigerian situation commenced on multiple culture contacts..., involving variant historical points in time as well; such that it appears rather more appropriate to talk about diversity and not necessarily “diversification” from Nigerian English foundations (Ugorji 2015:38).

While I find the effort to locate the place of NigE in the dynamic model particularly persuasive and interventory, its major deficit, perhaps, is the absence of synchronic data to prop the clarifications and counter-positions raised in Ugorji’s response. Similar to earlier comments on the variety, he relies largely on gleanings from reports which themselves were premised on poor empirical evidence (cf. Olajide & Olaniyi 2013, Olaniyi 2014). Besides, such instantiatory flaws as discussed in Ugorji would be unwarranted – in view Schneider’s call for ‘further testing’ of the parameters against global realities (Schneider 2003:273). It is thus fair to appreciate the theoretical concession to geolinguistic relativities, and the limits of its applicability. Despite various criticisms, the model has made significant theoretical contribution to our understanding of diachronic and synchronic development of English in new environments, and has been found to apply to most varieties of English around the world (Melchers & Shaw 2011).

Perhaps, more striking was an overarching post-appraisal in Schneider’s (2014) New reflections on the evolutionary dynamics of world Englishes. Beyond recounting of spoils, he admits the constraints the model has faced in some Outer Circle’s varieties (cf. Mesthrie & Bhatt 2008, Mukherjee & Gries 2009, Van Rooy & Terblanche 2010, Bekker 2009, Huber
reiterating that the limitations in the model ‘derives from the very nature of a model, which is an abstraction from reality, not reality “itself,” which highlights specific aspects disregarding others’ (Schneider 2014: 17). It’s true that the model was initially designed without the Expanding Circles in mind, but the need for their inclusion is becoming expedient (Kachru & Nelson 2006, Davydova 2012).

Table 4.1: The applicability of constituent components drawn from the Dynamic Model in Expanding Circle countries and emergent contexts. The ‘+’ sign indicates application to the category, ‘−’ suggests it doesn’t; ‘?’ marks uncertainty; ‘( )’ parentheses imply weak validity to the category; ‘/’ slashes indicates some degree of uncertainty between the two categorisations. Abbreviation: China (English), Korean (English), Japanese (English), (English) in ASEAN, Poststructuralist use (from Schneider 2014: 27)

As an extension, Schneider has in a subsequent analysis annexed a sub-model (Table 4.1), which attempts to account for realities in some expanding ecologies such as: China, South Korea, Japan, Thailand, Namibia, and Rwanda; ascribing the sociolinguistics of
English in this circle to what he terms ‘Poststructuralist diffusion’. The countries represent territories where the forms and use of English are ‘marked by the availability and adoption of (bits and pieces of) English in all kinds of creative and innovative contexts, to varying extents’, and the functional exigency of the language is prioritised among the speakers (Schneider 2014:25). In essence, the nuances of PCEs, and those of the Inner Circles pose a complex conundrum for theoretical proposals, however hybrid – a dilemma Schneider has, perhaps, also admitted as too dynamic for the Dynamic Model. Instead of a paradigmatic sketch (as in the Dynamic Model), a different conceptualisation transnational attraction\textsuperscript{15} is proposed. ‘Attraction’, in this sense, evolves on the concept of an ‘attractor’, a loaned parallel from the chaos theory. The motivation is mainly utilitarian, and revolves around the functional tendencies of English as an attractor on the global scale, and as a versatile medium for speakers of varying nativities constantly exposed to structural approximations and diverse social factors.

My elaboration of theoretical clusters and thoughts on the evolutionary realities of English in this section hold a range of connections with the crux of this dissertation. While my intention has not primarily focused on the re-assessment of theoretical stances highlighted above, I propose a further engagement based on data from such varieties for which the models were originally conceived, and from whose observations stronger conclusions can emerge – hence, making the goal both attractive and onerous. For instance, it would be practically unlikely that a study builds an honest corpus – enough to counter-sign or challenge some of Schnieder’s abstractions, considering the swathes of variables that would need to be included.

Basically, a proper sociophonetic investigation of the Dynamic Model, for instance, would entail gathering a varied pool of data from the territories concerned – a task too ambitious for the current study (even within the Nigeria alone), and for which this study is remotely unintended. It is likewise reasonable that only studies with enough data representation on English as spoken by over 500 native tongues in the country can pull such a feat. As much as these hold, an argument could also be made on the basis of a phenomenon some NigE scholars have either described as ‘convergence’ or ‘convergence of educated usages’ (Bamgbose 1995, Banjo 1995, Ugorji 2015). The notion presupposes a range of overarching typicality or ‘consistent patterns of structural and non-structural properties of the language which are taken to be typical’ (Ugorji 2015:36). It therefore follows that measured generalisations may surface or be made, albeit moderately, from the investigation of a sub-

\textsuperscript{15} Schneider maintains that since the current state of English is roundly transnational, and weighs much beyond the bounds of the Concentric Circle (Kachru, 1992); only an abstraction which captures the ‘transnational turn’ being experienced in user nations around the world suffices.
variety, as with data from a small subset, e.g., Ebira English in the wider contexts of Nigerian Englishes.

### 4.1.6 Conclusion

The foregoing has, very generally, drawn eclectically from an array of theoretical fronts I consider adaptable to my research questions. For example, very relevant to the overall concept of L2 phonology (especially in terms of lectal variation) is the selective potential of idiolects and dialects from feature pools and the tendency towards imperfect replication of original items from the pools (Mufwene 2002, 2003). The synergy of Mufwene’s idea of variationist possibilities in linguistic ecologies and Schneider’s evolutionary stages for post-colonial Englishes both underscore the speakers’ selective potential from the linguistic facilities available to them as well as the conditioning influences of their social universe. As earlier explained, though the descriptive reach of Schneider’s PCEs would be sufficient for some foremost L2 systems, the social realities of NigE suggest differently – just as Schneider himself has noted in one of his post-appraisals (Schneider 2003:241). This, most likely, warranted the lone reaction so far – by Ugorji (2015) as well as in Schneider (2014) in Section 4.1.4.1 above. For instance, in view of obvious variationist trends in NigE accents, even from the outset, the suitability of phonological ‘convergence’ or the ‘convergence of educated usages’ has been proposed (Bamgbose 1995, Banjo 1995, Ugorji 2015), thus reinforcing the key social parameter often implicated in reports on NigE accents, i.e., speakers’ education level or degree.

The following section broadly expounds on major normalisation procedures, and considerations made with regard to the choice of a particular method prior to statistical analysis. First, I summarise some key normalisation goals and their specific importance to current study. As will be seen shortly, none of the methods, especially their underlying algorithm is completely without flaws. It thus becomes expedient to highlight the computing strengths and weaknesses of some of these methods – towards applying the most relevant of them.
4.2 Data Normalisation

Normalisation procedures rank particularly crucial to sociophonetic studies in that they allow, to a substantial measure, cross-speaker comparisons of formants values that reflect actual sociolinguistically relevant information while excluding artifactual or extraneous variations. Since speakers’ formants could differ depending on their vocal configurations – often due to biological factors such as age and sex, vowel normalisation is therefore necessary for standard cross-assessment of speakers with different acoustic tendencies (Thomas 2011:160). The importance of normalisation, perhaps, became profound following Peterson & Barney’s (1952) analysis of tokens from 76 General American speakers, consisting of men, women and children. They observe that the first formant for the children is about half an octave higher than those of the men, and the second and third formants were equally higher; such that ‘variation of the measured data for a group of speakers is much larger than the variation encountered in repetitions with the same speaker’ (Peterson & Barney 1952: 182-3).

Speakers’ vocal tract sizes are inversely proportional to formant frequencies, in the sense that a small vocal tract would produce high formant frequencies and vice versa. For instance, children, and most females often produce higher formant frequencies than adult males due to their short vocal tracts – which makes it difficult to determine whether the differences in the acoustic properties are socially relevant or are just mere consequences of physiological differences. It is unlikely that a vowel retains an exact formant value for different speakers, and that a speaker would yield uniform formants for the same vowel, even in similar phonological contexts (Peterson & Barney 1952: 182, Hofmann 2015:199).

Despite the lack of accord over which normalisation algorithm is generally most suitable, raw non-normalised Hertz frequencies from different speakers would be less ideal for direct comparison (Watt & Fabricius 2002, 2009, Flynn 2011, Brato 2012; Hofmann 2015). Thus, normalisation basically helps to: i. minimise or eliminate variations caused by physiological differences among speakers; ii. preserve inter-speaker variation due to social category differences, including age, gender and dialect, or due to sound change; iii. preserve phonological distinction among vowels; iv. model the cognitive processes that allow human listeners to normalise vowels uttered by different speakers (Thomas 2002, 2011, Langstrof 2006, Thomas & Kendall 2007; Disner 2008, Fabricius Watt & Johnson 2009; Clopper 2009, Flynn 2011, Brato 2012, Hofmann 2015).

Though all the goals are variedly important to different shades of acoustic studies, scholars in sociophonetics are generally keener on the first three goals. The fourth goal has so
far been dispensable for research in traditional sociolinguistics, but could also be highly influential in sundry instances of language change and cognitive processing (Fabricius, Watt & Johnson 2009: 415, Thomas 2011: 161). Most available algorithms, understandably, function to accomplish some of these goals. While it is usually tricky to find an overly ambitious method that fulfils all the goals; the burden of preferences ultimately rests on the researcher – in view of the objectives at hand (Thomas 2002, 2011, Flynn, 2011). Since majority of the available normalisation techniques were originally built for specific studies, and assessed based on how successful they performed with the data, none can be overarchingly optimal. However, some of them are quite powerful in the reduction of artifactual variation (Thomas, 2002; Hofmann, 2015).

Depending on the configurations, the plethora of available techniques for normalisation draws on two major concepts, i.e. the \textit{vowel-intrinsic} and \textit{vowel-extrinsic}, and \textit{speaker-intrinsic} or \textit{extrinsic} methods. The categorisation thus depends on whether a corresponding algorithm takes just the individual vowels or the entire classes of vowels that are present in the variety into account. The vowel-intrinsic normalisation relies on the laundry of acoustic information extracted from a single vowel; e.g., taking the F1, F2, F3 & F0, amplitude, bandwidth or duration of one formant to normalise the formant without taking other neighbouring vowels into account (Syrdal & Gopal 1986), while the vowel-extrinsic method factors in the formant values of vowels spoken by the same speakers. Another approach is whether to base the normalisation on a speaker’s features or from the entire population sample (cf. Labov 2006, Thomas & Kendall 2007), but apart from Labov’s (2006) successful instantiation of this model on an American variety, the speaker-extrinsic normalisation is not as popular and often considered counter-productive (Flynn 2011). With speaker-intrinsic methods, mean computations are relative to each speaker, thus including just a single layer of normalisation in the algorithm; while the speaker-extrinsic approach effectively considers the grand mean for all speakers in the study.

Apart from Syrdal & Gopal’s (1986) ranking of the intrinsic over the extrinsic methods – in modelling speech perception, recent comparative studies have reckoned the overwhelming success of vowel-extrinsic methods (Adank et al 2004, Flynn 2011, Thomas 2011). As with the speaker-extrinsic methods, vowel-extrinsic procedures are most optimal with more vowel classes than with just a subset of speaker’s inventory. Both the vowel- and speaker extrinsic methods can come readily handy if, for instance, the study focuses on just one or very few number of classes and a large number of speakers. The vowel-intrinsic
method is equally considered immune to dissimilarities in the phonological inventories of dialect and languages, hence effective for interethnic or cross-dialectal comparison (Thomas 2011:165). The implementation of the vowel-intrinsic methods however depends on either the *fundamental frequency* or the F3 for normalisation – whose measurements (especially for sociolinguistic purposes) often lack acoustic integrity. Also, rhoticised vowels predominantly have lowering effect in F3, making the comparison of vowels in pre-/r/ contexts with others difficult. Other possibilities range from the mouth’s size to creakiness or nasality in the speaker’s voice, all of which may also skew the F3 (Thomas 2011:165).

The vowel-intrinsic normalisation however enjoys much attraction in sociophonetics research. In fact, the list of 20 normalisation methods examined on space equalisation and alignment in Flynn (2011) has 12 vowel-extrinsic, 7 vowel-intrinsic, and the raw Hertz values. While each of these algorithms is significantly effective in its own right, I discuss details of the extrinsic methods proposed in Lobanov (1971), Nearey (1978), Watt & Fabricius (2002), Gerstman (1968), which have been weighed as some of the most effective and straightforward procedures for sociophonetic data (Fabricius et al 2009, Thomas 2011, Flynn 2011, Brato 2012, Hofmann 2015).

Figure 4.2: Idealised vowel triangle for the construction of the centroid S. i=min F1, max F2; a=max F1; u’=min F1, min F2, where F1 (u’) and F2 (u’) =F1 (i) (Watt & Fabricius 2002:164)
Of these methods, the Watt & Fabricius (hence W&F) is most evolved and varied. In its oldest variant, the speaker’s vowel space is constructed as a triangle, with the edges representing the minimum and the maximum F1 and F2 values – by which the grand mean (centroid) is calculated (Figure 4.2). The point of the FLEECE vowel [i] accounts for the maximum values of the speaker’s F1&F2, while the [a] coordinates are the sum value for the TRAP/START tokens, also representing the speaker’s maximum F1&F2 values. Thus, [u'] is constructed such that F1[u'] = F2[u'] = F1[i], indicating the maximum F1 and F2 for the speaker (Watt & Fabricius 2002:163, Flynn 2011:5). The centroid for normalisation is then derived from the mean values of the coordinates: [i], [a] & [u'] on the edges of the triangle, being hypothetical representation of the most frontish, lowest and backmost points of the vowel space, based on this algorithm (Watt & Fabricius 2002):

$$S(F_i) = \frac{F_i[i] + F_i[a] + F_i[u']}{3}$$

The model has however been assessed inadequate for failing to account diligently for the dynamism of low vowels, especially across more than one varieties. It is true that some varieties (e.g., most African English varieties) may actually approximate their low vowels to the space represented in the schema; some inventories of North America tend to show a ‘butterfly pattern’ with two vowels at the very bottom of the envelope (Thomas & Kendall 2007). In response to this, the F2 value of [a] in the calculation of S(F2) is disregarded in the modified version (mW&F); such that calculation is made only on the basis of the distance between F2 of [i] and F2 of [u'], and not three F2 values as in the original formula (Fabricius, et al 2009:421). Another adaptation is in Kamata (2008) in which KIT and START are used due to the unique quality of FLEECE and TRAP in the studied variety.

$$S(F_i) = \begin{cases} \frac{F_i[i] + F_i[a] + F_i[u']}{3}, & i = 1 \\ \frac{F_i[i] + F_i[u']}{2}, & i = 2 \end{cases}$$

Gerstman’s (1968) method is another rather powerful vowel-extrinsic normalisation procedure but less prominent relative to other techniques, despite strong recommendations on excellent performance (cf. Adank 2003, Adank et al 2004, Clopper 2009, Flynn 2011). In a multi-conceptual comparison of different procedures, Gerstman’s ranks second to Bigham’s
(2008) on the hierarchy of performance in terms of formants equalisation and alignment within the vowel space for different speakers (Flynn 2011:23). Its low prominence is, perhaps, understandably due to age (being the oldest of earliest methods), and non-inclusion among available techniques in NORM suite (See Thomas et al 2007, Kendall et al 2010). The Gerstman’s algorithm lines up the two frequency extremes of the speaker’s vowel space, between the scale of 0 and 999:

$$F_{Gerstman}^{ti} = 999 \times \frac{F_{ti} - F_{ti}^{min}}{F_{ti}^{max} - F_{ti}^{min}}$$

where $F_{ti}^{min}$ is the minimum value of $F_i$ for all nine vowels for speaker $t$ and $F_{ti}^{max}$ as the maximum of $F_i$ (Adank et al 2004: 3101). One of the earliest vowel-extrinsic normalisation techniques was introduced in Lobanov (1971). Basically, it normalises formant values by subtracting a speaker’s mean formant value from all vowel tokens, before dividing them by the standard deviation (Adank, et al 2004, Flynn 2011). Aside from Disner (1980) who notes that all the vowel-extrinsic methods as poorly performed at preserving the ‘linguistic validity’ of the data compared to the vowel intrinsic procedures, Lobanov has been rated as among the best in several comparisons with other normalisation procedures (Nearey 1977, Adank et al 2004, Fabricius et al 2009; Flynn 2011). Its potential has, perhaps, also been reinforced by its capability to yield normalised values that translate into readable F1/F2 plots and the straightforwardness of the underlying algorithm:

$$F_{n[v]}_{n} = \left( F_{n[v]} - MEAN_{n} \right)/S_{n}$$

$F_{n[v]}_{n}$ is the normalised value for formant $n$ of vowel $V$, while $MEAN_{n}$ is the speaker’s mean for formant $n$ and $S_{n}$ as the standard deviation for $n$. The disadvantages with the model, however, are those shared by all vowel-extrinsic methods, which is that they tend to be most effective when all the speaker’s vowels are included. And similarly, it would be unfit for cross-dialectal or linguistic data. Building on Lobanov, four additional methods: Nearey1, Nearey2 or NeareyGM & two exponentials of Nearey have also been in wide use – two of which (formant-intrinsic and formant-extrinsic methods) are explained here (Nearey 1978). Both compare the speakers’ vowel space by aligning them at their mean formant frequencies. The first Nearey1 deploys a range factor for each formant, while the other Nearey2 shared logmean makes use of similar range across the formants (Adank, et al., 2004, Clopper 2009, Flynn 2011). Though the methods have been adjudged as generally impressive and mostly superior (Adank, et al., 2004: 3105, Thomas 2011: 166); all formulations of Nearey’s fare
comparatively worst in the multi-conceptual tests of 20 different methods, while Nearey1 was found to have performed slightly better than other variants (Flynn 2011:22). The two formulae are as follows:

\[ F_i^N = \ln(F_i) - \mu_{1_{n(F_i)}} (\text{Nearey1}) \quad \& \quad F_i^n = \ln(F_i) - \left(\frac{\mu_{\ln(F1)} + \mu_{\ln(F2)}}{2}\right) \]

Choosing a normalisation technique for my data was not as easy, since there had not been previous studies on the variety in which any of the methods was used. Thus, I relied on the literature, i.e. on previous comparisons of the procedures, and some of the initial outcomes from implementing two of the best models on the tokens: Lobanov and W&F’s methods. In the comparison of his procedure with Gerstman’s (1968), Lobanov reports his method to have performed best at mitigating the spread of vowel points spoken by different speakers and in managing the distances between adjacent dissimilar phonemes. It is equally observed to have behaved most excellently well in filtering out physiologically-dependent variations in formant values while retaining sociolinguistic differences (Adank, et al., 2004). The revised version mW&F is generally similar to Lobanov and Nearey in terms of potential. Fabricius, et al. (2009), from the evaluation of performance in achieving optimal conditions for visual representations, confirms Lobanov as most successful, while W&F performs ‘nearly as well, and in some cases better than Nearey’s CLIH₂’ (Fabricius et al., 2009:431). Their results further show the three models as well performed in preserving angles calculated against the F1 dimension and conclude that the S-centroid W&F procedures perform at least as well as the two most recognised speaker-intrinsic, vowel-extrinsic procedures. Beyond seemingly self-assessments in Lobanov (1971) and Fabricius et al (2009), an independent evaluation by Clopper (2009) which tests Nearey’s procedures with other methods finds Lobanov’s, W&F’s and Gerstman’s algorithms to have performed best among the total seven compared. His conclusions, however, can be read with a pinch of salt, considering that data from only two speakers were used in the analysis (Flynn 2011: 8). The trio of Lobanov, Nearey1 and Gerstman also excel significantly in accounting for vowel perception, coming solidly ahead all vowel-intrinsic procedures. In a further comparison among 20 different methods in Flynn (2011), the cluster of Gerstman’s, Nearey, ImW&F and Nearey are found to have performed well in equalising the vowel spaces – with Gerstman outperforming them all.

Fabricius et al (2009:431) have however advised the importance of inductive testing of the data sets and the purpose of study while deciding which procedures would be most suitable for normalisation. Thus, prior to selection, I delimited my options to just two,
namely: *originalW&F*’s and Lobanov’s methods – because of their relatively fair performances in the literature. A multi-way ANOVA was used to weigh the differential effects for KIT/FLEECE with gender and age as predictors, along with the interaction between the variables. Both show very similar effects – with Lobanov-normalised values producing a much stronger effect for gender; and *originalW&F* for age. For normalising the Ebira English data, consideration was given to Lobanov for its ease of implementation, fittingness to the current data, and excellence in the preservation of sociolinguistic variation in previous comparisons. In a range of discriminant analysis in Adank et al (2004:3102-3), Lobanov comes well ahead of Nearey1 and Gerstman in preserving phonemic variation and reducing anatomical/physiological differences, and equally turns out to be the best procedure in preserving key shades of sociolinguistic distinctions among the speakers. Notwithstanding the widely shared concerns about the risk of exposing salient elements of sociolinguistic information to acoustic levelling (cf. Flynn 2011:2), the Lobanov procedure, followed by Nearey1 and Gerstman have been found to retain most sociolinguistic variation in the normalised data (Labov 2001, Adank et al 2004).

### 4.2.1 Statistical Assumptions and Modelling

In the absence of statistically-developed theories on NigE vowel patterns, the need to rummage the data for much clearer grasp of possible trajectories became expedient. As discussed in Chapter 2, my key research questions were necessarily derived from prior auditory investigations in the literature, hence the need for a bottom-up approach that formed constituted my pilot findings. The bottom-up approach is mostly considered ideal for sociophonetic corpus since it entails the exploration of a large, previously uncharted corpus for hypotheses that might best explain the patterns. Unlike the top-down procedure, the bottom-up designs enjoy wider prominence in sociolinguistics because of their emphasis on fieldwork, survey frameworks and naturalistic techniques of collecting data (Thomas 2011:2). The statistical tool for data evaluation was, for the most part, determined by key considerations – which I briefly underscore below, as well as further illustrations on the models.

To begin with, since there is rarely an overarching method for exploring acoustic representation of mergers or near-mergers, the use of at least two different metrics is recommended (Nycz & Hall-Lew 2014). Thus, graphical representations of vowels on the two-dimensional plots, and basic diagnostic evaluations helped much in pinpointing some of
the ‘supposedly’ coalesced vowels for statistical tests at the later stage. Preliminary plotting was done using an R adaptation of a vowel manipulation and normalisation package (Kendall Tyler & Erik 2012). By default, the package executes some normalisation procedures, including Lobanov, Watt&Fabricius, Nearey (Section 4.2), and has a range of built-in pre-fab functions for generating vowel plots in the F1/F2 dimensional space. For distributional overview of vowel pairs, boxplots were additionally generated to reflect the quantiles and largest observations. Boxplots are relatively honest in quantifying pair distributions with the median scores and interquartile percentiles. With boxplots, the spread of each crossed vowel with factors and their levels are more visually evident, and outlying tokens outside the whiskers in each distribution distinct are rendered obvious. The variance in the whiskers’ length shows much about the normality and the skewness of the distribution, and the length of the box points at the kurtosis of data points near the median (Hofmann 2015:211).

4.2.2 Linear Regression Models

Among tons of statistical methods for analysing naturalistic data, the multiple regressions seem the most suitable, given its capacity to evaluate the simultaneous effects of several contextual predictors on the outcome variable (Johnson 2014:7). In most sociophonetic data, for instance, response variables (i.e., F1/F2) are often analysed with linear regressions. The method is basically predictive in explaining the relationship between a dependent variable and one or more explanatory factors. An example of how it works would be an attempt at determining the role of gender in pitch behaviour for a population. In that case, the biological sex of speakers is fixed as the main predictor of voice pitch in Hz.

Beneath the linear regression model is the general linear algorithm\textsuperscript{16} that takes in a continuous or numeric dependent variable and predictor(s) with an error. It equally permits the inclusion of random effects (Gorman & Johnson 2013:231). Like many other statistical methods, linear regression assumes a range of conditions (assumptions) the dataset must fulfil for its results to be taken seriously. These include, among others, linearity, homoscedasticity, normality, absence of collinearity and independence (Winter 2013, Hofmann 2015, Griess

\textsuperscript{16} The general linear equation for linear regression is $Y = b_0 + \Sigma(b_iX_i) + \epsilon$ where $Y$ represents the continuous dependent variable and the usually continuous or binary predictors as $X_i$, and $\epsilon$ as the error representing the variance that is not accounted for in the model. To denote inter-dependence of the response variable, the addition of $j$ to $Y$ and the general error $\epsilon$ would yield the following modification: $Yj = b_0 + \Sigma(b_iX_i) + \epsilon j$. 
2009). Also in practice, regressions hardly include every important predictor, neither are the predictors ever perfectly uncorrelated — thus rendering its results grossly suspicious, especially with sociolinguistic data (Johnson 2014:8). Linear correlation assumes strong causative effects on dependence measures between the response and predictor(s) variables.

In R, linearity can be assessed on residuals plot by graphing the standardized residuals on the y-axis and the standardised predicted values on the x-axis. Homoscedasticity presupposes that the measures of variance across the data are equal for the predicted values, such that the variance of the errors does not depend on the predictors (Winter 2013:16, Johnson 2014:8). Tests for linearity tend to work better with continuous variables as response and predictor(s). Attempt to plot some of my numerical/categorical predictors such as speakers’ chronological age or the age of exposure to L2 against F1&F2 yielded a bunch of striped graphs, thus unfit for assessing linearity. The pair of F1/F2 can, nevertheless, be weighed against each other to determine the relationships between both dimensions. Again, this would not hold for all the lexical sets, in view of differences (in advancement and retraction) between vowels’ configurations. For instance, while we expect a low and high F1 & F2 frequencies (respectively) for high front vowels, other vowel classes deviate from the linear effects of either variable. As evident in the following plots, Figure 4.3 indicates a mild

Figure 4.3: F1/F2 residual plots of FLEECE&KIT showing mild violation of linearity assumption
violation of linearity assumption on the F1&F2 residuals plot for FLEECE & KIT (both high front vowels). With formant values, it is important to also evaluate their distributions (Hay 2011:204). Normally distributed data (also known as Gaussian distribution) are symmetric, unimodal, and asymptotic, and the mean, median, and mode are all equal. They tend to spread symmetrically around the mean value, such that each side of the bell-shaped frequency curve equals the other. Parametric procedures including correlation, t-tests and ANOVAs assume intrinsically that the samples drawn are normally distributed, and as such derive their statistical validity on this assumption (Ghasemi & Zahediasl 2012:486).

Similar to other statistical assumptions discussed below, the test of normality requires some attention. Depending on the type or size of data, fulfilling or violating normality could have serious implications on the validity of statistical conclusions (Asghar & Zahediasl 2012). With regard to phonetic outcomes, most variables are rarely distributed normally (Hay 2011:200). Due to the nature of continuous variables in sociophonetic data, their non-normality is neither consequent on extreme outliers, nor does a relatively ‘cleaned’ dataset guarantee a normal spread. The basic requirements are seldom met by naturalistic corpus in any way. For example, pitch and formant frequencies are hardly ever normal. Continuous dependent measures such as vowel formants (F1 or F2), and variables such as the chronological age across sociolinguistic groups and lop-sided numeric predictors are ‘normally’ prone to normality violations. The random effects of words, speakers and the influential effects of phonological co-texts on formant values can drive the deviation of sample quantiles from the theoretical quantiles – against which the actual distribution is plotted. Often, acoustic data are either skewed to the right or left, i.e., with single tail of distribution noticeably longer or stretched out to one side than to the other tail. A ‘skew’ could be positive or negative – depending on the density concentration.

However, a large amount of tokens (typical of sociophonetic samples) are generally immune to the statistical assumption of non-normality violation, hence the equal fittingness of parametric procedures for testing such data. For instance, an approximate sample size of > 30 and above would mitigate normality issues, even with parametric procedures (Pallant 2007). Linear and mixed effect regressions are equally robustly resilient against the normality violations (Winter 2013:18). Nonetheless, non-parametric procedures with less stringent assumptions are considered mostly fit for analysing non-normal distributions (Hay 2011). Aside assumptions about randomness and sample independence, non-parametric tests are robust against probability distributions and outliers. But their statistical power are weak in
that, compared to parametric tools, larger amount of data are required to achieve significance – which is somewhat desirable for analytical integrity (Hay 2011:201).

For exploration and testing of tentative hypotheses, I used scatterplots, the Norm Q-Q plot (also quantile-quantile plot) function, boxplots, and histograms for visual assessments. The Norm Q-Q plot mainly reflects the similarity of two datasets distributions by comparing the sample distributions against a normal distribution and plotting their quantiles against each other. Usually, the points in the Q-Q plot would dot on the line $y = x$ to show that the pair distributions being compared are alike. Statistical tests for normality are additional to visual representations (Elliott & Woodward 2007). Basically, the sample scores (sample quantiles) are compared to a normally distributed set of scores (theoretical quantiles) which have the same mean and standard deviation, with a null hypothesis that assumes a normal distribution and the alternative hypothesis indicating a non-normal distribution – in the event of statistical significance. With such test, small amount of data are, however, characteristically weak in rejecting the null hypothesis (normality threshold) while large samples are sensitive to high probability even in cases of minimal deviation from normality (Field 2009:822).

The F1 distribution for KIT&FLEECE in Figure 4.4 & 4.5 appears to violate the normality assumption. Though the histogram is somewhat bell-shaped and the Q-Q plot indicates some degree of parallel on the straight line, the Shapiro-Wilk test of correlation between the data and the normal scores was significant ($w = 0.96367$, $p$-value = 2.2e-16). The Shapiro-Wilk test is a conservative procedure with a strong sensitivity to outliers which tends to reject the null hypothesis in favour of an alternative hypothesis in the event of non-normal distribution. Thus, the high $p$-value for the data indicates an outright violation of normality in the dataset. With such dataset, a less stringent way of assessing the variance, especially between the mean values of hypothetically merged vowels would be the Kruskal-Wallis rank test. It is an extension of Two-sample Wilcoxon test for more than two levels of independent variables in single analysis. The latter is usually run as a post-hoc test to detect the exact level(s) of significance among all tested predictors.

$$H = \frac{12}{n(n+1)} \sum_{i}^{k} \frac{R_{i}^{2}}{n_{i}} - 3(n + 1)$$

The statistic estimate is based on a chi-square distribution with $k-1$ degrees of freedom if the null hypothesis of equal distribution of frequencies is true. The validity of its approximation assumes each of the $n_{i}$ to be at least 5 in count. Unlike parametric procedures, both
Figure 4.4: Normal Q-Q plot for normalised F1 of FLEECE&KIT (1904 observations, 29 speakers; across all categories)

Figure 4.5: Normal Q-Q plot for F1 normalised values of KIT (1904 observations, 28 speakers across all categories)
parameters are non-parametric, hence distribution-free and far less stringent in dealing with outliers. Nonetheless, Kruskal Wallis is not entirely assumption free. It assumes randomly drawn samples from the source population, and independence within and between variables (Andale 2016). The dependent variables must be ordinal, ratio or interval scaled, and all groups should favour the same distributional patterns, i.e., uniformly tailed/skewed towards either a negative or positive distribution. Another is the independence of each dependent variable (i.e., each observation), which is by far the most crucial in linear tests, which if violated, could yield spurious statistical significance, i.e., Type 1 error (Hofmann 2015: 215).

Most non-parametric procedures assume that each of the responses actually come from different subjects – such that each speaker would have only contributed a token of each vowel class to the formant pool. A classic example would be the flip of a coin or the roll of a die, each of which is completely independent from the outcome of the other (Winter 2013:21). In the case of a phonetic data, independence assumes the absence of intra-speaker variation across the token inputs, i.e., the absolute independence of each outcome of continuous variables as in F1/F2. The question of independence is more or less a choice between the devil and the sea! In actual fact, non-independence is much likely true of most linguistic data, and similarly an issue in many branches of science, yet this doesn’t seem to prevent linguists from using the tests (Lazic, 2009, Hay 2011, Winter 2011, 2014). Fortunately, resolving non-independence in statistical analysis and forestalling the chances of errors is a question of the statistical model, and this is where the mixed models become invaluably convenient (Winter 2013:21). Apart from mixed models, there is seldom any route around non-violation of independence. However, the mixed effects model (used in this study) is robust against the independence assumption – with variables such as the individual speaker and word items held as random effects (Section 5.2).

An equally important problem to grapple with in a regression model is multicollinearity, especially in dealing with multiple predictors. Some two or more predictors would be collinear if they are more highly correlated with other co-predictors or independent variables than with the outcome variables. Since multiple regressions involve more than one variable – each of which is assumed independent of the other, undue collinearity can complicate or compromise the optimal variables. For instance, in a data set with the prospect of age and job status or education having a collinear effect on certain vowel realisation; chances of recovering the specific predictor of influence becomes blurred, as each of the fixed effects could effectively strip the other of predictive potential. In other words, the presence of
highly correlated predictors in a model automatically complicates the determination of which of those actually playing prominent roles (Winter 2013: 15).

Collinearity would not be much of a problem in a regression equation if the goal is to predict the overarching effect of a cluster of predictors on a dependent variable – in which case the regression would still predict the overall $R^2$ or still reflect the predictors’ random and fixed effects, as well as the goodness of fit of prediction. However, it poses fundamental concerns when the goal is weighing the contributory effects of specific variables to the outcome, as in the case of this study. In such cases, the odds of either Type I or Type II error are higher and the regression coefficients’ confidence intervals could become erratically sensitive to slight adjustments in the data set. Dealing with collinearity is therefore important, and can be done in a number of ways. A rather straightforward approach involves pre-empting the problem at the design stage of the study by simply leaving out obviously collinear variable(s) from the model ‘just to see what happens’ (Winter 2013, Williams 2015). These solutions however are not without serious setbacks. A sociophonetic data as in the current study are habitually designed from the outset with certain predictors in mind. Excluding such variables for others in the model could impact seriously on the research questions. Also, if the variable really belongs in the model, this may result in specification error, which can be even worse than multicollinearity (Williams 2015: 4).

4.2.2.1 Choosing Predictors for Statistical Modelling

Since my study had very little of theoretical standpoints to lean on with respect to the most interesting variables, all possibly important explanatory predictors had to be included in the regression – bearing in mind the odds of including those with similar predictive potential. Though this was attractive, the need to exclude those that might have collinear effects from the final list of predictors was even more important. For diagnostic evaluations, therefore, I ran separate linear regressions with the all predictors in my dataset, in order to assess the level of intercorrelations between them using Variable Inflation Factor (VIF). Calculating the VIF is one of the most basic ways of detecting collinearity among multiple predictors. A VIF value for a single explanatory variable is obtained using the $r$-squared value of the regression of that variable against all other explanatory variables:

$$VIF_j = \frac{1}{1 - R_j^2}$$
where the VIF for variable j is the reciprocal of the inverse of $R^2$ from the regression. For instance, in the search for predictors with collinear effects on the status of FLEECE and KIT in F1 & F2, I dumped all variables into one model. While a relatively low VIF for predictors suggests absence of multicollinearity, inclusion of those with comparatively higher values needed further review for relevance and uniqueness.

<table>
<thead>
<tr>
<th>f1_normalised variables</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>66.3</td>
</tr>
<tr>
<td>age_at_exposure</td>
<td>3.0</td>
</tr>
<tr>
<td>yrs_of_exposure</td>
<td>55.6</td>
</tr>
<tr>
<td>age_group</td>
<td>12.1</td>
</tr>
<tr>
<td>gender</td>
<td>1.2</td>
</tr>
<tr>
<td>edu.degree</td>
<td>1.3</td>
</tr>
<tr>
<td>No.of.L1</td>
<td>1.9</td>
</tr>
<tr>
<td>job.status</td>
<td>11.6</td>
</tr>
<tr>
<td>style</td>
<td>1.0</td>
</tr>
<tr>
<td>phone_label</td>
<td>1.0</td>
</tr>
<tr>
<td>phone_duration_n</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 4.2: Estimation of collinearity among predictors for FLEECE/KIT in F1 & F2: n = 18740; speakers = 29

The definitive scale for ‘high’ VIF values is somewhat subjective, but values above the borderline of 10 are commonest. Thus, in Table 4.2, predictors in bold fonts – such as: chronological age, years of exposure to English, age-group and job status correlate strongly in their effects on FLEECE/KIT realisation. Reviewing the variables in question, correlation between job status and age-group was plausible, due to the fact that all young speakers were mostly university/polytechnic students (except one who was already in the civil service as at the time of recording), while the older generation were mostly civil servants. Similarly, strong correlation was conceivable between speakers’ chronological age and the years of exposure – as the older speakers, logically, would have been exposed to English far earlier and much longer than their younger counterparts. Based on this diagnostics, I re-assessed the following independent variables against their VIF: chronological age=66.31, years of exposure to English =55.57, age group =12.1 and job status = 11.6.

Perhaps, an additional grand method of detecting collinear predictors is via classification and regression trees (CART), also known as Decision Trees. The technique has become very attractive for exploring predictors’ hierarchical importance – as well as hinting
on variables’ significance in the data set. It could also reveal predictors with collinear effects on the dependent variable in regression models (Mendoza-Denton, Hay & Jannedy, 2003; Hay, 2011). CART is a non-parametric method of regression trees which embeds tree-structured regression models into a well-defined premise of provisional inference procedures. Decision Trees basically partition the data into different nodes using the same response variable. They start by sorting out the most influential predictor in relation to the dependent variable before recursively partitioning the data into clusters of significance nodes until it is all taken into account (Breiman 2001, Eddington 2010, Hay 2011). In the case of a sociophonetic dataset, for instance, if the formants of a vowel are entered as response variables alongside phonological contexts and other predictors, the tree would effectively separate the formant mean values according to the phonetic environments or the social factors in question (Hofmann 2015:209). This what is done for gender in Figure 4.6.

Considering the regressive potential of CART algorithm, its predictive accuracy is usually adjudged as similar to VARBUL. In comparison, while CART could make generalisations with the logistic regression model, Decision Trees can provide more coarse-grained output (Eddington 2010:265). However, while the paucity of recursive representations as customary to CART, classifications from the tree are often useful in data mining and exploration, and results from it may be reported alongside or as alternative analysis for the data (Eddington 2010:282-3, Hay 2011:213).
Figure 4.6: Decision Trees of predictors for FLEECE/KIT_F1, showing the initial splits for gender.
An enhanced procedure for CART is Random Forests (Breiman et al. 1984, Breiman 2001), which generates a large set of trees using different subsets of data. Random Forest does not over-fit and is very effective in isolating outliers, thus paving way for multiple inclusion of variables (Eddington 2010, Hay 2011). Recursive partitioning can be implemented in the *party* package in R (Hothorn et al. 2006, 2015). The package equally implements recursive partitioning for survival data and its main function, i.e., *ctree* or *randomForest* can take in continuous, nominal and multivariate dependent outcomes in a conditional inference framework. Each tree was followed up by post-hoc analysis of variable importance and corresponding varImp plots (Figure 4.7). On the whole, the functions make use of the so-called out-of-bag (OOB) samples to conduct prediction accuracy. While the varImpPlot(s) assists very accurately in outputting the list of predictors that are most important to the response variable (i.e., F1 or F2 of the Lexical Sets), the *importance* function further provides a pair of informative %IncMSE (Mean-square error) and IncNodePurity (node purity) that show variables’ significance. The most revealing of both measures is the %IncMSE which reflects a predictive increase after the variables are permuted or their values are shuffled randomly. The higher the number of %IncMSE, the more important it is in the model. In addition, the varImPlot can also hint on collinear effects of predictors within the model (i.e., independent variables whose absence or removal from the model portend no harm).
Figure 4.7: VarImp (Variable importance) plots for FLEEECE/KIT_F1 with the highest prediction for gender: %IncMSE: FLEEECE = 55.94, KIT = 45.32 & IncNodePurity: FLEEECE = 7.91, KIT = 7.46 (see Appendix B.1–B4) for other varImpPlots (GOOSE/FOOT, LOT/THOUGHT and TRAP/BATH).
4.3 Methods for ‘Mergers’

A long-established method for determining the status of merged vowels is based on the Euclidean Distance (ED), also known as Cartesian or Pythagorean Distances between their mean values in F1 & F2 dimensional space (see Labov, Ash & Boberg 2006, Baranowski 2007, Nycz & Hall-Lew 2014). Essentially, the distance between the vowel classes is constructed as the hypotenuse of a right-angle triangle, with the remaining sides as F1 & F2. The Pythagorean Theorem is therefore deployed in measuring the length of the hypotenuse – such that a smaller value suggests a tilt towards a merger while a greater value indicates differentiation. For measurement of vowel formants, it is the squared root of the sum of two quantiles, i.e., the squared difference between the mean values for the vowel classes in both dimensions $ED = \sqrt{(F2 − F2)^2 + (F1 − F1)^2}$. Apparently, the underlying procedure for calculating ED is straightforward and practically easy, as it can be applied to formant values without recourse to more sophisticated statistical alternatives. Also, since the only variables that can be assessed with this methods are the raw vowel Hertz, its results tend to be lot handy and much readable in terms of distance length.

In addition, a Euclidean metric accounts for both F1 & F2 at the same time, thus inclusive of both dimensions in one swoop. Nevertheless, ED poses a number of shortfalls, especially for weighing naturalistic tokens. Most importantly, ED deems the relative contribution of F1 and F2 to the distance measure as equal, thus ignoring the possibility of imbalances between the acoustic measures or their variance (Gorman & Johnson 2013:232). The measurement is also less statistically sound since it actually never indicates any significance value; as well as ineffective in accounting for other cluster of predictors (such as: phonological/co-textual environments, demographic factors, etc) which are very influential in determining the vowel trajectories. A sounder substitute for ED is the multivariate analysis of variables (MANOVAs). The method, which improves on ANOVA, can be used to fit a multivariate outcome in which the lexical sets being assessed (i.e., the phonemes) are the main predictor with F1 & F2 as dependent variables. As per estimating the status of merger, the most important information is the statistic of MANOVA output known as Pillai score (see Hay, Warren and Drager 2006, Hall-Lew 2009, Wong & Hall-Lew 2014). The Pillai score ranges from 0 to 1 relative to the length of distance between the vowels. While 0 may indicate a loss of distinction, higher scores suggest phonemic differentiation. The score in this sense represents the amount of the multivariate variance accounted for by the vowel class predictor
Though both ED and MANOVA can concurrently accommodate two dependent variables in a model, the latter is more advanced in that other acoustic parameters (such as: amplitude, duration, etc) can be included alongside into the model (Di Paolo 1992, Nycz & Hall-Lew 2014), and the ability to factor the correlative nature of the two dependent variables into the analysis (Hofmann 2015:229).

Undesirably, calculating Pillai scores requires the equality of covariance and the sample size, failure of which might result in Type 1 error. And as would be expected in linear-based tests, MANOVA assumes independence of samples, and is much likely to yield spurious results with small tokens (Hofmann 2015:230). For the historically merged/non-differentiated vowel classes in my data, results from the linear regression should be statistically significant for the phonemes in F1&F2, especially after the random effects of word, speakers and the phonological contexts had been factored into the model. In events this was the case only in either of the dimensions, a MANOVA (with both F1&F2 as dependent variable) was triangulated with linear regression outputs.

One flexible method of mitigating the shortcomings of the foregoing, especially those that apply in the case of ED is ED-Adjusted or Adjusted Euclidean Distance (Nycz & Hall-Lew 2014:4-5). The procedure is designed as a follow-up to the results of mixed effects regression models. Mixed effects regression is notably robust against normality and other crucial statistical assumptions against which most parametric tests are weak. Theoretically, a mixed model can include as many predictors as there are to be tested, and random variables such as the effects of individual word and speaker – therefore accounting for independence of observations and the sway effects of phonological co-texts (see Section 4.4.2 below for review of this model). However, the tool can only estimate differences or significance in either of F1 or F2 in a run, and not simultaneously. The thinking is that the coefficients generated from the regression results can thus be used to compute the ED-Adjusted. Again, though the method remarkably improves on EDs based on the raw mean values, the information as to whether the difference(s) between the vowels are statistically significant or not remains lost.
4.3.1 Mixed Effects Modelling

In this section, I explain the statistical mechanisms of mixed effect modelling and its aptness for my analysis. My selection of the method was informed by its computational possibilities and resilience; and growing popularity for assessing naturalistic linguistic data. For the assessment of linguistic variations, the tide has widely shifted towards mixed effects modelling for estimating the effect sizes between response and predictor variables (cf. Baayen, Tweedie & Schreuder 2001, Baayen, Davidson & Bates 2008, Gorman 2010, Gorman & Johnson 2013). Fixed effect modelling has invariably performed very weakly in factoring into the analysis – the range of complex factors which often equally impact most linguistic responses being examined. One of the major setbacks of the fixed effects models is its powerlessness in handling of ‘nested’ predictors, especially at lower-levels, i.e., failure in quantitative estimation of variables such as word items and speakers (Johnson 2014:1, Nycz & Hall-Lew 2014:4). A simple regression model often represented as follows: $Y_i = \beta_0 + \beta_1 X_i + (e_i)$ basically assumes a fixed effect for predictors, and independence of observation from the sampled population (Quené & Bergh 2008:414). The random term $(e_i)$ in the formula is the residual that marks the aspects of the response or dependent variable not explained by the model. Thus, $X$ represents the range of predictor vectors, which indicates that $\beta_0 + \beta_1$ are overarchingly fixed in the model.

A major weakness of ordinary regression, perhaps, is the inherent assumption of independence of observation. This, however, reinforces the usefulness of the mixed effect model – which improves on the within-group correlation often present in grouped data (Johnson 2009:364, Winter 2013). So far, the most reliable way of relaxing the homogeneity assumption is by including random effects, thus accounting for speakers’ idiosyncrasies in the analyses (Johnson 2009:377). The requisites of homogeneity, in essence, are reasonably beyond naturalistic speech data, i.e., casual speech. Factors such as speakers are often drawn from a larger population, and tokens for a particular lexical set (in the case of sociophonetic data) are drawn from different grammatical words and phonological contexts. While speakers are usually classified into groups of social categories, for instance, phonemes are mostly identified with their corresponding lexical sets. In mixed effects model, dependencies such as these are however accommodated by adding a list of random variables so as to explain the random effects (Quene’ & Bergh 2008:414). Thus, the mixed effects modelling has much lower risk of chance probabilities arising from fixedness of predictors’ effects, and has

In linear fixed effects modelling, phenomena are rigidly conceived as ‘what is doing what’ with less consideration for miscellaneous factors (internal and external) that are equally influential, thereby heightening the risk of Type 1 errors. Usually, this form of modelling has a general error term $\varepsilon$ added to the list of predictors on the right hand side of the dependent variable in the model which accounts for the ‘probabilistic’ or ‘stochastic’ sides of the model (Winter 2013:2). For example, a hypothetical ordinary regression model for the assessment of predictors’ effects on the first formant (F1) in my dataset would be:

$$f1_{n} \sim \text{speaker}_\text{age} + \text{age}_\text{of exposure} + \text{edu. level} + \text{style} + \text{gender} + \text{job status} + \text{phone} – \text{label} + \varepsilon$$

The model combines 7 separate predictors with an error term $\varepsilon$. But the functionality of this seemingly complex fit ignores the idiolectal idiosyncrasies of speakers and word items that could have equal influence on the predictors’ effects. More importantly, it ignores the fact the pool of observations in the data are not independent (i.e., each speaker actually has more than one token for each lexical set in the dataset). Multiple responses from a speaker cannot be independent of each other nor can different speakers classed into a social category share exactly the same mean values across the vowel sets. This reality, unavoidably, would render a fixed models’ output inter-dependent rather than dependent (Winter 2013:2).

In addition, the simple regression model without random variables is weak in mitigating the sway impacts of outliers on the effects. In mixed effects regression, individual speakers and word or items vary in the model as ‘random effects’ – as such lending more validity to inter-group effects (Sigley 2003, Baayen, Davidson & Bates 2008, Johnson 2014). The mixed effects models method has elegant measures against errors that could result from individual speaker’s idiosyncrasies and item-induced effects. They raise the precision of the regression results by ‘partialing out’ or accommodating these random factors, such that they do not misrepresent the regression coefficient of main predictors, but are backgrounded to the variance explained by these intervening effects, i.e., items and speakers (Gries 2015:97). Each factor of a random effect gets a coefficient, thereby providing a way for the factors to vary from one another (Drager & Hay 2012: 61). The consequence of such outcomes includes the odds of specious significance extraneously triggered by random variables confounded along the main predictors.
As noted in Johnson (2009:369 & 2014:23), the inescapable concession in mixed models is their equal proneness to Type II error, ‘where a real population difference does not show up clearly or consistently enough in the sample to be recognised as statistically significant’. A large amount of data and a good number of subjects can however minimise the likelihood of such errors (Johnson, 2009). The distributions of sociophonetic corpus are usually uneven, mostly noisy and messy across levels such as styles; and for speakers or groups to be compared. Also, researchers often hardly succeed in recruiting an equal number of speakers for each cell of the social groups being studied. The fixed effect regression, regrettably, is less tolerant of these imbalances, but averages each speaker/group and the tokens equally across the population. My design, for instance, contain four social groups into which speakers were split before data collection. But I inevitably had to return with unequal number of speakers for each group (which would mean unequal data distribution across speakers’ group). Running a simple regression model on the data could yield spurious results that are primarily due to these imbalances, not genuine social effects; hence the more fittingness of mixed effects modelling for my data.

4.3.2 Random Intercept in Mixed Models and By-speaker Analyses

Aside the diligence of accounting for inter-group significance by suppressing the within-speaker and word effects in mixed effects modelling is a painful compromise of discarding individual speaker variation – especially when the effect of the variable being analysed vary from speaker to speaker. Recent proposals in Drager & Hay (2012) and Johnson (2014) saliently draw attention to the importance of low-level variations – mostly occurring at idiolectal layers. Model fitting with speakers as random variables essentially mutes possible within-speaker variations in favour of between speakers’ difference, thereby treating random effects intercepts as by-products of the model (Drager & Hay 2012:59). They argue that apart from the standardising effect of random intercepts in mixed models, they can also shed more understanding on individual speaker’s style and identity. Since each speaker’s behaviour with variables can be uniquely captured in a multilevel modelling, interpreting their separate adjustments to the regression coefficient would allow for fine-grained inroads into the bed of dissimilarities at the idiolectal level (Section 4.1.1). Accessing individual speaker’s baseline in relation to their group thus provides us with richer awareness of likely innovators or conservatives in the overall course of phonemic mutation.
Random intercepts account for within-group differences not factored into the main significance effects. In mixed effects regression, overarching intercepts predict the expected behaviour for the infinite population - based on the available data, while the random intercepts indicate the measure of divergence from the general intercepts. Because random intercepts are summed based on the model’s prediction, separate tendencies across the random intercepts of multiple speakers can be read as unique deviation from the fitted outcome, i.e., the individual speaker’s adjustment to the regression coefficient that reflects the pattern in which they consistently differ from the others (Griess 2015:97). Drager & Hay (2012:62) suggest that in interpreting random intercepts, both directions as to whether the intercept is positive or negative in values can be resourceful. They explain that while a speaker with a positive intercept is prone to pander towards the variant being modelled (e.g., retracting, advancing, lowering or raising of a vowel variant), those with negative intercepts are seldom to use it, and that the closer the intercept is to 0, the more conservative the speaker is in relation to the variant. I find the exploitation of random intercepts for assessing the gradience of vocalic patterns at individual speaker’s level mainly instructive – which fortunately, is also part of what Rbrul’s provides.

4.3.3 Rbrul for Mixed Effects Regression

Rbrul is front-end interface to R environment designed as a radical enhancement to GoldVarb and other statistical software for logistic regression (see Rbrul version 2.3.2: Johnson 2016). Similar to its precursor, it is most viable for multiple logistic regressions – where the probability of one outcome is modelled as a linear outcome of multiple predictors. Other than Rbrul, the commonest implementation of the variable rule programme VARBRUL is GoldVarb X (Sankoff et al 2005, Brato 2012). The variable rule models were designed to assess the effects of multiple factors on binary linguistic levels – the presence or absence of a response, or any phenomenon treated as an alternation between two variants (Johnson 2009: 359). By implementation, the software identifies those factors (linguistic or social) that have influential effects on the regressor variable, and to what extent such can be quantified. GoldVarb remains a famous tool of choice in quantitative sociolinguistics, despite inherent statistical shortcomings. While Rbrul is elegant with both continuous and categorical regressors, GoldVarb suffers a restriction to logistic regression, thus requires only categorical data as response variables. A logistic regression requires a discrete data as response variable,
e.g. the variants [θ], [t], [f] of a consonant variable such as /θ/ or gender, marital status, job as independent variable. It cannot deal with continuous vocalic properties (F1/F2) predominant in most sociophonetic data (Brato 2012:77).

The non-versatility of GoldVarb, especially its restriction to binary variables, and the manner of presenting regression outputs renders it unfit for diverse kinds of evaluations, and limits its interactions with other disciplines outside sociolinguistics (Johnson 2009:360). For instance, a factor with different levels is known as factor groups in GoldVarb. Outputs for factor effects are called sum contrasts, where each coefficient suggests a deviation from the mean; and treatment contrasts, where one level of each factor or variable is set at a coefficient of 0. Each of the other levels is then assigned a coefficient representing the effect on the response of switching from the baseline to the ‘treatment’ level in question. Also, coefficients are expressed in log-odds either in positive or negative numbers – instead of probabilities ranging from 0 to 1. Log-odds are obtained based on \( \ln[p/(1-p)] \) where a positive value moves towards the effect, and a negative value suggests retraction from the effect, and 0 is without effects (Johnson 2009:361, Brato 2012:78). While most variable rule software report their outputs either in log-odds or factor weights, Rbrul harmonises the dichotomy by presenting results in both formats, thus making interpretations more accessible and allowing sociolinguists to better relate with studies in cognate fields.

Though GoldVarb still enjoys some prominence among sociolinguists, the platform is innately sluggish, constrained and likely to produce anti-conservative results based on spurious factors. On the other hand, Rbrul runs much faster on personal computers, and is excellent in handling both continuous numeric response variables and predictors – which means that the effects of continuous predictors such as speakers’ chronological age, age of exposure to L2 and phoneme duration can be simultaneously modelled with vowel formant values (F1 or F2). The software is quite supple with handling interactions between two or more predictors having collinear effects, and its significance threshold of 0.05 can either be manually adjusted or by the Bonferroni correction which splits \( \alpha \) by the number of independent variables, such that with 8 IV, the \( \alpha \) is set to 0.05/8 \( \approx \)0.006 (Field 2009, Brato 2012). Depending on the fitted model, Rbrul can report \( R^{2} \) or pseudo \( R^{2} \) as part of the outputs; and is more resilient with regard to the number of factors and levels that can be included in each model.

Despite its many frills, the choice of Rbrul is a measured trade-off between statistical efficiency and some shortfalls too. Rbrul, as currently run in R, can neither test
assumptions for the model being built nor detect multicollinearity among predictors. It also
excludes the standard errors of coefficient in its output. However, the VIF scores, which I
took for all predictors against each dependent variable (Figure 4.7) and CART (Section
4.2.2.1), mainly helped to make up for this particular limitation. Also, Rbrul is more
prone to Type II error, as it is less likely to report factor effects than other software. In a simulated data
used to test the effect of gender on the realisation of [n] as [ing] or [n] in Rbrul and GoldVarb,
the Type II error representation for Rbrul becomes higher than GoldVarb’s as the individual
variation increases – thus unable to reach significance for underlying gender effect (Johnson,
2009):

The trade-off between Type I and Type II error is an ever-present issue in statistical
analysis, and has no simple solution. Most researchers would probably endorse a
conservative approach, arguing that it is better to overlook something that does exist
than to report something that does not. This attitude would lead us to prefer Rbrul
(Johnson, 2009:369).

4.3.4 Interpreting Rbrul Outputs

Unlike VARBRUL and GoldVarb – which run multiple logistic regression only with binary
or categorical response variables that represent discrete linguistic alternatives, Rbrul combines
this with linear regression with the possibility of continuous response variables (such as
vowel formants: F1 or F2), and can simultaneously estimate the effects of continuous
predictors on these variables (Johnson 2009:363, 2014:7, Hofmann 2015:233). Bearing in
mind that formant values are continuous, linear regression is certainly the favoured analysis
for my modelling. Rbrul reports different forms of outputs for different types of regression.
Below, I discuss what some of these statistical outputs generally mean about the configuration
of the models and the factors entered.

With a binary response as in logistic regression, Rbrul models the natural logarithm of
the odds of the response, thus yielding the log-odds \( \ln(p/(1 - p)) \), where \( p \) is the odd of the
response or a linear function of the predictor. Log-odds are sourced from the natural logarithm
of the odds, such that the odds are the probability of the certain linguistic behaviour (in the
case of a linguistic data), divided by the probability of it not occurring (Johnson 2009:361,
2014:8). While a positive value suggests a tilt in favour of the effects, negative value
disfavours it; and the value of 0 is neutral. Another outcome for logistic regression is factor
weight, which is redundant, i.e., it reports the same thing with log-odds. Factor weight represents probabilities ranging from 0 to 1. However, both representations can be uniquely instructive – such that the while log-odds may show a more accurate fit of each category to the data, factor weights are helpful in drawing comparison across different sets of data (Johnson 2009).

The representation of outputs is however different with linear regression. Instead of log-odds and factor weights, Rbrul outputs the raw coefficient and mean values for each factor in the predictor. Put simply, regression coefficient parallels the mean change in the dependent variable for each unit of change. It indicates the units of difference from the Y-intercept (the predicted value of X on the y-axis when Xs are set to zero) as a result of change in a particular X while other Xs are held constant. It expresses a statistical control whereby the effect of one factor is isolated from the rest in the model and weighted. Each factor variable is thus assigned a regression coefficient. Unlike with continuous predictors, a unit of difference between a categorical predictor such as age in my dataset would mean a switch from 0 to 1 (i.e., from the category of the younger YNGR to older OLDR speaker group). Thus, the coefficients would represent the mean disparity in Y between the reference category YNGR = 0 and the comparison category OLDR group = 1. In my analysis, however, the coefficients for such binary factors ranged from negative to positive, thus expressing measures of between-group differences. Similar to outputs for logistic regressions, the raw coefficients and the corresponding mean values basically report the same information, but in different forms. However, it is important that the coefficient for each factor correlates with the mean value, i.e., factors with high or low coefficients should yield corresponding mean values. Factors with lower coefficients but higher mean values could be indicating interaction, and may need to be removed from the model. In event that such was due to small amount of tokens in my data, I manually merged the factor in question with similar factor(s) to boost its number of tokens.

In addition to the α levels (P-values) for fixed predictors, Rbrul also reports measures such as Akaike’s Information Criterion (AIC) and the $R^2$-squared ($R^2$). Both reflect the involvedness or suitability of the fitted model from different standpoints. While the AIC returns the penalized measures for the predictors’ relevance and the number of subjects in the model (Hofmann 2015:236), $R^2$ explains how well the model fits the data, or how much the actual data deviates from the prediction of the model. In case of AIC, their values may increase as additional predictors are added, but decrease if the added predictors better
accounts for the observed variance in the response variable – such that while a larger $R^2$ indicate a worse the fit, a model with lower $R^2$ is more suitable. However, a very low $R^2$ fixed value often suggests the poor effects of main predictors (fixed effects) on the variances explained, a relatively high value indicate the prominent impacts of the model. Therefore, a model with a low $R^2$ fixed and high $R^2$ random values strongly suggest a more of idiosyncratic effects principally occasioned by the speakers and word items rather than the sociolinguistic categories (predictors) in entered into the run. Traditionally, while a higher range of $R^2$ fixed value from 0.20 (20%) to 0.50 (50%) is often convenient in the analysis of native varieties, a value between 0.15 (15%) and 0.20 (20%) would be reasonable for L2 varieties (Hofmann, p.c.). Unlike AIC, the $R^2$ values tend to remain unchanged when more predictors are added to the model (Hofmann 2015). Though the interpretation of what AIC values say about the model is usually subjective, those with comparatively smaller values are assumed to be much better fits than those with larger values.

The application of the mixed effect modelling in the measurement of some of the historically attested non-differentiated vowel pairs in the NigE systems is presented in Chapter 5. As discussed briefly in Section 1.1, the methods employed in the statistical assessment equally draw on some of the procedural conventions for analysing vowel mergers. For instance, vowels in each pair are classified as phonemically distinct or otherwise, depending on the significance, i.e., $p$-values returned for phone label in the model. Though separate models were built for the analysis of vowels in both formants, and in the preceding, following and combined phonological co-texts/contexts, their fittingness for each analysis was determined mostly by comparing their $R^2$ values and the AIC’s.
5 Analysis of EEng Monophthongs

5.1 Overview

The analysis in this chapter mainly describes the monophthongal inventory of EEng in the light of previous notions on the vocalic structure of NigE. Since a model system for EEng was yet unclear, I adapted Wells’ (1982) Lexical Sets for the Southern British English (SBE) for coding the vowels into vocalic classes. The intent, however, was not to contrast my findings with the SBE, but as basic reference for the taxonomy of English vowels. While there is some agreement on the homophonous status of some vowels in NigE, studies are conceivably divergent on the scale of occurrence and the effects of linguistic factors. For example, Josiah & Sola (2011:544), in a diachronic review of proposals for the phonemic catalogue of NigE recount a number of disparities. Based on the studies, the number of monophthongs is pegged as follows: 5 (Christopherson 1954), 11 (Eka 1985, Odumuh 1987), 7 (Adetugbo 2004), 12 (Ekong 1978, Jibril 1991), 7 (Awonusi 2004), 6 (Eka & Udofot 1996, Udofot 2004), 8 (Jowitt 1991) and 7 in basilectal, 8 in mesolectal & 12acrolectal varieties (Ugorji 2010). Notwithstanding the disagreements, some consensus on vowel coalescence and contrast in NigE has been consistent. The perennial loss of contrast between the tense/lax pairs of FLEECE & KIT, FOOT & GOOSE, LOT & THOUGHT and BATH & letter, including the monophthongal status of FACE and GOAT have been variedly attested (Ekong 1978, Banjo 1995, Eka & Udofot 1996, Adetugbo 2004, Udofot 2004). In a rather interesting turn, the prospect of differentiation between these vowels – commonly by ethnicity and speakers’ level of formal education – has also been etched on some of the findings. An instance of such exceptions is the odds of distinction between FLEECE & KIT, BATH, TRAP & letter as well as FOOT & GOOSE in Educated Hausa English (Jibril 1982, Odumah 1987, Jowitt 1991). As explained in Section 2.4, though there are yet to be strong empirical explanations for these variations, Awonusi (2004) & Gut (2004) have fingered the influence of the Southern British English (SBE) education on the earliest rank of Hausa elites, as a result of their contact with RP-speaking British school teachers, even after Nigeria’s independence in 1960. Consequently, the pair of high front KIT [i] & FLEECE [i:], FOOT [u] & GOOSE [u:] and the cluster of TRAP[æ], BATH [ɑː]
& lettER [ə] have been marked as differentiated in Hausa English (Jowitt 1991, Jibril 1995). Aside claims of substrate influences from the L1 systems, no similar logic has however been put forward for the identical overlap of these vowels in Southern Nigerian accents.

Figure 5.1 is a synopsis of vocalic behaviours in EEng. Basically, it presents a general illustration of historically fused vowels in the NigE system. The graph, however, requires a cautious reading, especially bearing in mind the odds of tokens imbalances between the classes (Nycz & Hall-Lew 2014:4) as well as other sundry conditions not factored into data visualisation at this stage. Nonetheless, they serve as basis for an impressionist overview of speakers’ vocalic configuration in the system.

![EEng vowels acc.to lexical sets](image-url)

Figure 5.1: Normalised F1/F2 plot of EEng vowel system (including diphthongs) for all speakers = 29; (n= 18585)

At first glance, the vowels in Figure 5.1 appear to form a trapezial ring of clusters round the envelope. There are relatively thicker concentrations of vowels in the high front and back and low central regions, with isolated occurrence of NURSE and DRESS through the mid–central areas. With the exception of superficial proximity of phonemes around the peripheral space,
the illustration signals a number of contrasts to the long-held notion of a five-vowel system in NigE. From the plot, tokens of the so-called j-words (Mesthrie 2010:13-16) classified as USE in this study are visibly frontish, i.e., they maintain a marked distance from the high back clusters. Cumulatively, the nurse [ɜː] vowel appears located in a central position, with noticeably lower F1 than the low central group and higher F2 than the low back groups (cf. Eka 2004, Adetugbo 2004, Simo Bobda 2000 & 2007). Among diphthongs, the nucleus of goat [əʊ] reflects a noticeably lower F1 than the low back vowels, while that of cure [ʊə] is also distinguished from this group (cf. Eka 2004, Adetugbo 2004, Simo Bobda 2007). Though the offsets of square and near show a uniform in-glide movement, their nuclei are considerably distinct in the mid-front area, therefore suggesting differentiation between the two classes and a manifest separation of their nucleus from the low central trap/bath (cf. Adetugbo 2004, Simo Bobda 2007:414).

Though the analysis of the diphthongs and other variants of major classes such as force, cloth, north, palm and start are beyond the scope of this study, the snapshot gives a refreshing overview of wider trajectories of vowel realisation in the system, and more importantly, provides a general basis of appraisal of previous findings. For instance, as a result of the cluster in the low back region and the obvious absence of rhoticity in the system, the trio of force, cloth and north were hence disregarded, retaining just lot and thought in analysis. Similarly, the palm and start sets were left out of the low central group, remaining trap and bath. The prospect of exploring the status of happY vowels against the high front group was hindered by extreme token imbalances between them (happY =137) as against (kit =1082 & fleece =881). Subsequently, kit and fleece were retained in the high front position. For a similar reason, all comma tokens (n=24) were excluded, leaving letter (n=373) for comparison with the low central vowels (Section 5.2.5.1).

Figure 5.2 shows only vowels conservatively reckoned as homophonous in NigE – on which this study is focused. Again, at a first view, the plot demonstrates some consistency with the literature. Though in close proximity, phonemic separation is indicated between the high front kit & fleece in F1 & F2, as well between goose and foot in F2 (see also Jibril 1982 & Odumuh 1987 on Educated Hausa English). This impression however attracts some inkling, particularly in view of goose dispersion in F2 – which reasonably, could have triggered a superficial mean difference between the sets. The coalescence of low back lot & thought, theoretically, is inter-ethnically complete in NigE (Banjo 1995, Adetugbo 2004, Udofo 2004, Awonusi 2004, and others) while strut might be achieved as separate phoneme
by some educated speakers (Udofot 2004 & Eka 2004). A single system has been reported for the sets of TRAP, BATH & lettER in major West African varieties (Udofot 1990, Gut 2002 & Akinjobi 2006, Section 2.5.2), quantitative differentiation between these vowels has also been noted in Educated Hausa English accent – a claim ostensibly reinforced by the nearness of TRAP & BATH in Figure 5.2. Also, in a virtual agreement with previous accounts of Hausa English, lettER is lowered in F1 and apparently moved away from the low central vowels (see also Jibril 1982, 1995, Jowitt 1991). Although these graphical appraisals suggest certain routes to our understanding of the system, they are impressionist, and would require further statistical evaluations to determine their underlying status.

Figure 5.2: F1/F2 plot of the historically merged sets in NigE (speakers = 29), comprising tokens from wordlist, reading passage & sociolinguistic interview (n=8004)

The following section begins with details of procedural considerations in fitting regression models for the runs. Before analysis, vowel sets were distributed according to consonantal environments in the following and preceding co-texts. And considering the higher number of factors for phonological contexts/co-texts – than other predictors in my data, a fair
distribution of tokens among the levels was necessary. The illustrations of how this was done for the high front and back vowels are in Table 5.2 – 4 and 5.7.

The procedures for the analysis of KIT and FLEECE were more complicated than for other pairs with theoretical standards of assessment. Unlike for back vowels, the incidence of high front merger has so far been sparsely-mentioned in studies on established English varieties. Consequently, it was initially unclear as to the dimension (F1 or F2) or the phonological co-text (preceding or following) in which to run a regression for the high front classes (cf. Gorman & Johnson 2013). Given the conundrum, I ran separate analysis in both dimensions, and for the following, preceding and both phonological co-texts. Though significant values were yielded for phone label in all the statistical runs (which suggests a differentiation between the classes), only results for both co-texts (hence complex model) in F1 was most representative – based on the AIC values (Table 5.6). Fitting procedures for GOOSE and FOOT commenced in Section 5.1.2. In grouping the high back sets, the fronting effect of /j/ (also known as the ‘j-words’) largely frontish tokens of GOOSE was first considered (Mesthrie 2010 & Hoffmann 2011). From the outset, I had assigned tokens in this category into a separate class of USE vowels, and so, excluded from the pair of GOOSE and FOOT. Table 5.6 shows the tokens distribution of FOOT and GOOSE prior to statistical analysis. Generally, back vowels are more robustly distinguished in F2 (Gorman & Johnson 2013, Hofmann 2015), hence the only dimension of assessment for all back vowels in this study. For FOOT and GOOSE, the effect of phone label was significant in all the models (Table 5.7). While this effect may be underlyingly marginal, phonological co-texts appeared to have mainly triggered the differences between the two classes (Figure 5.4).

For low back vowels, the analysis of LOT and THOUGHT preceded STRUT. In all the models, results for LOT and THOUGHT suggest an absolute single system, although further measurement of STRUT against the low back group showed a minor differentiation in terms of vowel quality (Table 5.10 & Figure 5.6). The regression outcomes for the low central cluster were not as clear in the intial estimation. While the regression results in F1 showed the class of TRAP and BATH as a single vowel, the results indicated a differentiation in F2. Meanwhile, the visual impression in Figure 5.7 suggests a clear fusion between the classes. In the light of this complexity, a MANOVA based on Wilks’λ, was employed for comparison. The multivariate effect was insignificant for phone label, while the post-hoc ANOVA also returned no significant effect for phone label in neither F1 nor F2. Based on these outcomes, the singular status for TRAP and BATH was established. Defined as one phoneme, either of the
classes could hence be paired with lettER in subsequent analysis. Different from what was done for LOT/THOUGHT prior to comparison with STRUT; recoding TRAP/BATH into a single class before comparing it with lettER resulted in extreme imbalances between the two fresh pairs. To mitigate this, only BATH (n=778) was selected for comparison with lettER (n=401). And considering the robustness of F1 for the measurement of vowels in this region, BATH and lettER were assessed with formant one as the dependent variable. The effects of phone label and style were highly significant (Jibril 1986), Udofot 1990 and Jowitt 1991 on the distinctiveness of schwa /ə/ from the low central sets).

5.2 Model Fitting and Linear Regression Analysis

In this section, I explain the standard procedures followed in fitting linear regression models for the analysis of KIT/FLEECE, FOOT/GOOSE, LOT/THOUGHT/STRUT, NURSE and TRAP/BATH / lettER. Similar to previous illustrations, I show only in this section how the methods were employed in the analysis of the high front pair: KIT/FLEECE. The four vowel categories comprise those historically marked as non-differentiated in Educated Nigerian English (cf. Jibril 1982, Jowitt 1991, Simo Bobda 1995, Awonusi 2004). My preliminary objective, therefore, was to describe their status among EEng speakers, with the aim of defining the monophthongal inventory in the system. Further assessment of variation patterns for each of the vowels in the monophthongal system is in Section 5.3.1 (see Section 2 for Research Question 1 &2).

Since the prime goal was to assess differences or similarities between the vowel classes, I gave keen priority to token distributions in the consonantal contexts (or phonological co-texts). More than other variables, the effects of linguistic environments usually tend to be strongest on vowel behaviours, and equally resourceful in the determination of phonemic inventories (Nycz & Hall-Lew 2013:4, Hofmann 2015:228). To maximise the conditioning effect of the phonological co-texts, all tokens were cross-tabulated with respective co-texts (e.g., in labial nasal, nasal, fricative, obstruents, voiced, voiceless, etc.) in the preceding and following co-texts. This allowed for the close inspection of their distributions and the exclusion of factors/co-texts with relatively fewer tokens as well as those which occur in environments that might compromise the statistical evaluations. For example, apart from tokens in approximants contexts (which I already excluded from the dataset), those in preceding or following liquid contexts were also excluded from the models – due to their
lowering effects in F2 for front vowels (Ladefoged & Maddieson 1996, Di Paolo, Yaeger-Dror & Wassink 2011).

In distributing the tokens according to phonological co-texts, I adapted a model that comprised the key influential environments often implicated in the vocalic variations. Instead of the classifications according to place, manner, voice and larger groups in the following and preceding co-texts, Table 5.1 is much reduced, yet robust. Only a variable is coded for preceding co-texts, and 3 factors (place, manner & voice) for the following co-texts. Before cross-tabulation, the factors in the following environments were combined by interactions and re-assignments between them. The preceding factors – being in just a column – were straightforwardly cross-tabulated with their respective tokens.

For vowels without a specific dimension of measurement, e.g., KIT & FLEECE, separate regressions were run for each vowel pair. Consequently, the effect of phone label in preceding and following co-texts, and in F1 and F2 were separately examined. The complex model is a subsumation of the preceding and following co-texts in a single column, and so ensures the inclusion of all linguistic co-texts as one predictor in the model.

<table>
<thead>
<tr>
<th>preceding</th>
<th>following</th>
</tr>
</thead>
<tbody>
<tr>
<td>place</td>
<td>manner</td>
</tr>
<tr>
<td>labial nasal</td>
<td>apical</td>
</tr>
<tr>
<td>liquid</td>
<td>labio-dental</td>
</tr>
<tr>
<td>nasal apical</td>
<td>palatal</td>
</tr>
<tr>
<td>obstruent</td>
<td>velar</td>
</tr>
<tr>
<td>oral apical</td>
<td></td>
</tr>
<tr>
<td>velar</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Phonological classifications in preceding and following co-texts

The total tokens of FLEECE/KIT after initial cross-tabulation with the preceding contexts were (n=1963). FLEECE = 881 and KIT = 1082. Despite healthy distribution of tokens according to classes, further cross-tabulation of phonological co-texts with the vowels in Table 5.2 reveals a skew. For example, a disproportion between oral labial and oral apical and between obstruents and other factors indicate gross imbalances that could result in Type I or II error. Also, the 300 tokens contributed by liquids to the pool were excluded from the model due to
### Table 5.2: Initial cross-tabulation of all FLEECE & KIT tokens with preceding phonological contexts before adjustment (n=1963), and after adjustments (n=1573)

<table>
<thead>
<tr>
<th>preceding co-texts</th>
<th>FLEECE</th>
<th>KIT</th>
<th>total</th>
<th>adjusted distribution</th>
<th>FLEECE</th>
<th>KIT</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>labial nasal</td>
<td>33</td>
<td>26</td>
<td>59</td>
<td>labials</td>
<td>294</td>
<td>190</td>
<td>484</td>
</tr>
<tr>
<td>liquid</td>
<td>128</td>
<td>172</td>
<td>300</td>
<td>apicals</td>
<td>129</td>
<td>207</td>
<td>336</td>
</tr>
<tr>
<td>nasal apical</td>
<td>41</td>
<td>25</td>
<td>66</td>
<td>obstruents</td>
<td>240</td>
<td>385</td>
<td>625</td>
</tr>
<tr>
<td>obstruent</td>
<td>240</td>
<td>385</td>
<td>625</td>
<td>velars</td>
<td>12</td>
<td>116</td>
<td>128</td>
</tr>
<tr>
<td>oral apical</td>
<td>88</td>
<td>182</td>
<td>270</td>
<td>total</td>
<td>675</td>
<td>898</td>
<td>1573</td>
</tr>
<tr>
<td>oral labial</td>
<td>261</td>
<td>164</td>
<td>425</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pause</td>
<td>78</td>
<td>12</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velar</td>
<td>12</td>
<td>116</td>
<td>128</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>881</td>
<td>1082</td>
<td>1963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5.3: Initial cross-tabulation of all FLEECE & KIT tokens with following phonological contexts before adjustment (n=1963), and after adjustments (n=1875). The interaction between the three phonological parameters yielded 20 levels in the raw distribution. The adjusted distribution shows the outcome of cross-tabulation of the lexical sets with the co-texts after recoding and excluding certain levels in the raw distribution.

<table>
<thead>
<tr>
<th>following co-texts</th>
<th>FLEECE</th>
<th>KIT</th>
<th>total</th>
<th>adjusted distribution</th>
<th>FLEECE</th>
<th>KIT</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>apical.lateral.voiced</td>
<td>70</td>
<td>119</td>
<td>189</td>
<td>apicals</td>
<td>283</td>
<td>401</td>
<td>684</td>
</tr>
<tr>
<td>apical.nasal.voiced</td>
<td>66</td>
<td>91</td>
<td>157</td>
<td>labio-dentals</td>
<td>340</td>
<td>244</td>
<td>584</td>
</tr>
<tr>
<td>apical.stop.voiced</td>
<td>72</td>
<td>71</td>
<td>143</td>
<td>palatals</td>
<td>166</td>
<td>149</td>
<td>315</td>
</tr>
<tr>
<td>apical.stop.voiceless</td>
<td>75</td>
<td>120</td>
<td>195</td>
<td>velars</td>
<td>20</td>
<td>272</td>
<td>292</td>
</tr>
<tr>
<td>labiodentals.affricate.voiced</td>
<td>6</td>
<td>13</td>
<td>19</td>
<td>total</td>
<td>809</td>
<td>1066</td>
<td>1875</td>
</tr>
<tr>
<td>labio-dental.fricative.voiced</td>
<td>161</td>
<td>163</td>
<td>324</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dental.fricative.voiceless</td>
<td>9</td>
<td>29</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dental.liquid.voiced</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dental.nasal.voiced</td>
<td>25</td>
<td>9</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dental.stop.voiced</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dental.stop.voiceless</td>
<td>137</td>
<td>26</td>
<td>163</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatal.affricate.voiceless</td>
<td>36</td>
<td>5</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatal.fricative.voiced</td>
<td>64</td>
<td>29</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatal.fricative.voiceless</td>
<td>66</td>
<td>115</td>
<td>181</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatal.liquid.voiced</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pause.pause.pause</td>
<td>61</td>
<td>12</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velar.fricative.voiceless</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velar.nasal.voiced</td>
<td>4</td>
<td>220</td>
<td>224</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velar.stop.voiced</td>
<td>2</td>
<td>12</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velar.stop.voiceless</td>
<td>11</td>
<td>40</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>881</td>
<td>1082</td>
<td>1963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
its lowering effect in F2 for high front vowels (Ladefoged & Maddieson 1996, Di Paolo, Yaeger-Dror & Wassink 2011). Towards a ‘cleaner’ table, the number of factors after initial cross-tabulation was reduced by recoding and merging them with their respective places of articulation, while those preceded by pause and liquids were excluded. Thus, tokens in preceding oral labial and nasal labial environments were combined into one class of labials and oral apical & nasal apical as apicals. Those in preceding pause and liquid environments were also excluded (see Table 5.2). The table presents both raw and ‘cleaned’ distribution of tokens in the preceding co-texts.

Adjusting the distribution of tokens in the following environments was a bit trickier as it involved more than one column. First, I stringed place, manner and voice together in a column by creating interactions between them before cross-tabulation with phone label. The initial table was messy, consisting of 20 levels or factors. Cells with apparently skewed distribution were merged with others, and those which could not be ‘sandwiched’ into other levels were excluded. The choice therefore was to choose the particular parameter the re-assignment of tokens. For consistency, I again chose place of articulation and re-distributed the tokens according to occurrence before apical, labio-dental, palatal and velar in the following co-texts (see Table 5.4).

<table>
<thead>
<tr>
<th></th>
<th>FLEECE</th>
<th>KIT</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced.apicals</td>
<td>280</td>
<td>281</td>
<td>489</td>
</tr>
<tr>
<td>voiced.labio-dentals</td>
<td>199</td>
<td>190</td>
<td>389</td>
</tr>
<tr>
<td>voiced.palatals</td>
<td>70</td>
<td>32</td>
<td>102</td>
</tr>
<tr>
<td>voiced.velars</td>
<td>6</td>
<td>232</td>
<td>238</td>
</tr>
<tr>
<td>voiceless.apicals</td>
<td>75</td>
<td>120</td>
<td>195</td>
</tr>
<tr>
<td>voiceless.palatals</td>
<td>248</td>
<td>175</td>
<td>423</td>
</tr>
<tr>
<td>voiceless.velars</td>
<td>14</td>
<td>40</td>
<td>54</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>820</strong></td>
<td><strong>1070</strong></td>
<td><strong>1890</strong></td>
</tr>
</tbody>
</table>

Table 5.4: Adjusted distribution both co-texts (complex model). The initial levels were re-combined according to place of articulation and voice (n=1890). Before running the regression, voiced.velar (fleece = 6, kit = 232) was removed owing to extreme skewness between the lexical sets.

Subsequently, separate regressions were run with each of the co-texts, i.e., preceding, following or complex model – alongside other predictors including; gender, age group, normalised phone duration, education degree, style & the number of speakers’ L1 as fixed effects. The effects of speakers and words were held as random, in order to mitigate the
chances of spurious results that are extraneous to the fixed predictors. Either of F1 or F2 was entered as the dependent variable. As explained in 4.3.1, the nesting of random variables from the fixed effects is, by far, a major gain of mixed effects modelling over other regression models. Fixed effects predictors usually consist of few levels or factors which can be replicated in further studies, while the random effects of individual speakers or words (which are invariably drawn from a much larger population are hardly replicable (Johnson 2009:365, Baayen 2008:241). Based on these procedures, the phonemic status of each vowel pairs or groups was determined on the effect of phone label.

The raw distribution of token from the dataset obviously does validate the need for this procedure. For instance, running a regression with raw distributions as in Table 5.2 & 5.3 increases the likelihood of Type 1 errors and meaningless coefficients for factors as well as needless interactions between the levels of phonological classes (Johnson 2009, Gorman & Johnson 2013). While some factors have fairly equal number of both phonemes, e.g., in labio-dental,fricative, voiced & apical, stop, voiced co-texts, some others do not. One way of rectifying this without throwing away a large number of tokens was to merge the factors based on shared consonantal features. For example, tokens in following liquids and pause environments, and those which could not be combined with more robust factors due to phonological dissimilarities were also excluded. These done, phonological co-texts alongside other predictors were entered into the model – to assess the effect of phone label in the following environments.

Though co-textual influence often tends to be stronger in the following phonological contexts (Labov, Ash & Boberg 2006, Hofmann 2015:228), the effects of preceding co-texts can be equally important. The usual practice is to report either of both or note their separate effects on the vowels in the analysis. Another possibility would be to combine the preceding and the following co-texts in a single model, in a way that the phonological effects (of preceding and following co-texts) are embedded as one variable (herein ‘complex model’). As would be expected, the complex model, after cross-tabulation, became expanded into a massive total of 105 levels (factors), which were drastically reduced to a small number according to place of articulation and voice.

5.2.1 High Vowels: KIT & FLEECE

Based on the outlined procedures, each model was run on one-level analysis in Rbrul version 2.3.1 released on May 9th, 2016 (Johnson 2016). Phone label, age group, phonological
environments, gender, style, education degree, number of L1 and normalised phone duration (continuous) were entered as fixed effect predictors, with word label (lexical items) and speaker names (each speaker) as random effects. Since the most crucial dimension in the differentiation of KIT & FLEECE is theoretically unclear, statistical effects in both F1 & F2 were assessed in separate runs. Aside education and ethnicity, there are no prior assumption in NigE literature about other predictors with regard to high vowels. The model for following phonological environments was run with 1875 tokens on 14 degrees of freedom (Table 5.5).

In F1, the regression returned the strongest significance for the combined effects of co-texts ($R^2 = 0.10 \ p = 8.76e-06$), gender ($R^2 = 0.10 \ p = 0.0012$), phone label ($R^2 = 0.362 \ p = 0.00655$) and phone duration ($R^2 = 0.10 \ p = 0.0228$). Other factors included in the model were not statistically insignificant. The Akaike Information Criterion (AIC) was 1574 with $R^2.f\text{ixed}$ and $R^2.r\text{andom}$ at 0.10 and 0.31 respectively. The hierarchy of significance however differs in F2. The model showed the strongest significance for phone label ($R^2 = 0.10 \ p = 5.67e-05$), followed by phone duration ($R^2 = 0.10 \ p = 7.12e-05$), then co-texts ($R^2 = 0.10 \ p = 0.0029$) including style ($R^2 = 0.10 \ p = 0.00722$) and age group ($R^2 = 0.10 \ p = 0.0163$). The AIC in F2 was 1847 with the $R^2.f\text{ixed}$ and $R^2.r\text{andom}$ at 0.10 and 0.21 respectively. The effect of gender was however not significant in F2 for following co-texts ($R^2 = 0.10 \ p = 0.0682$). The effect of other Nigerian languages and education degree were also insignificant in F1/F2. Though the following phonological environments often tend to be more influential in vowel realisation (Clark, Yallop & Fletcher 2007, Thomas 2011, Brato 2012, Hofmann 2015), the preceding co-texts do equally contribute to the overall vocalic behaviour. Often, analysis in the preceding co-texts is ignored on the notion that their effects do ebb away before the point of measurement in either F1 or F2. While such decision depends on whether the formants were measured at mid-points or 50% (for monophthongs), or a little earlier within the steady state e.g., 33% or 25% and 33% (Evanini 2009: 65 & Hofmann 2015: 175), studies have shown both co-texts as equally very crucial in the assessment of vowel quality (Thomas 2011:49, Hillenbrand, Clark and Nearey 2001: 754, Hofmann 2015: 194).
Table 5.5: F1 regression results for KIT & FLEECE in following co-texts (n = 1875) and preceding co-texts (n = 1573). For following co-texts: $R^2$.fixed=0.10 $R^2$.random =0.27; and preceding co-texts: $R^2$.fixed=0.11 $R^2$.random =0.31. Only predictors with significant effects are shown on this table.
For preceding co-texts, speakers and word items were retained as random effects with same predictors as in the previous model. Gender led the significance order in F1 ($R^2=0.11$ $p=0.000773$), followed by phone label ($R^2=0.11$ $p=0.00437$). Other variables were insignificant. The $R^2$ for fixed effects yielded 0.11 and 0.31 for random effects. Deviance was -667 and the AIC value equalled 1424. In the second formant, however, phone label was most significant ($R^2 = 0.10$ $p=0.00156$). Also, phonological co-texts ($R^2=0.10$ $p=0.00249$), phone duration ($R^2 = 0.10$ $p=0.0061$), and style ($R^2 = 0.10$ $p=0.0115$) had significant effects. Age group ($R^2=0.10$ $p=[0.0978]$), gender ($R^2=0.10$ $p=0.141$), number of L1 ($R^2 = 0.10$ $p=[0.451]$) and education degree ($R^2 = 0.10$ $p=[0.716]$) were however insignificant. The $R^2$ values for fixed and random effects were 0.10 and 0.21 respectively. Deviance was slightly higher than in F1 at -799.492 and 1688.69 for AIC.

The purpose of different runs with the same predictors was to determine the best model in explaining the between and within-group variances – which is reflected in the fixed $R^2$ and Akaike Information Criterion (AIC) values. Despite slight difference in the results for following and preceding models in F1 & F2, phone label remained significant, especially in F2 where it came well above other predictors. Another prospect would be to include the same number of predictors in the complex model. For this model, the phonological co-texts showed the strongest effect in F1 ($R^2 = 0.11$ $p=1.61e-06$), and gender ($R^2 = 0.11$ $p=0.00092$), phone label ($R^2 = 0.11$ $p=0.0545$) with the deviance at – 505 and AIC = 1116. The $R^2$ values for fixed factors 0.10 and the value for random effects = 0.275. Effects were also significant in F2 for: phone label ($R^2= 0.10$ $p=2.92e-05$), and phone-duration ($R^2 = 0.10$ $p=0.000193$), phonological co-texts ($R^2=0.10$ $p=0.00254$), age group ($R^2 = 0.10$ $p=0.0303$), then style ($R^2 = 0.251$ $p=0.0389$) and gender ($R^2 = 0.10$ $p=0.0408$). Both number of L1 and education degree demonstrate weak effects according in this run ($R^2 = 0.10$ $p=0.567$) and ($R^2 = 0.10$ $p=0.844$). The deviance is – 755, AIC = 1615, and the $R^2$ value for random factors was 0.20 and the fixed factors’ was 0.10.

The two earlier runs for following and preceding co-texts both suggest KIT and FLEECE as separate phonemes. But the effects of phone label yielded by the complex model in F1 revealed a marginal significance ($R^2=0.11$ $p=0.0545$), but highly significant in F2 ($R^2 = 0.10$ $p=2.92e-05$). This discrepancy between the F1 and F2 results is, perhaps, explicable on the retention of tokens in following lateral environments in the data, thus resulting in KIT retraction in F2 (Thomas 2011:101).
By and large, a trend of differentiation between the KIT and FLEECE is demonstrated in the data. The next step therefore was to determine which of the models best explained the variances in the data, and in which dimension (i.e., either in F1 or F2). Usually, such hints are derivable from the Akaike Information Criterion (AIC), or the fixed $R^2$ values for each analysis. Though the AIC value has little potential in regarding the model’s goodness of fit, its comparison with values from other runs can be directly instructive (Hofmann 2015:237). The amount of explained variance in the model $R^2$ and deviance also can indicate a better performance. As a rule of thumb, a lower value for any of these parameters suggests a relative fitness of the model while higher values may be an indication of collinear relationship among the predictors entered into the model. It is also possible, on the other hand, a decrease in the $R^2$ values is possible when additional predictors are added. Generally, there appears to be no consensus in what $R^2$ reflect in the model (Nikagawa & Shielzeth 2013:134).

**following co-texts and preceding co-texts (complex model)**

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced.apicals</td>
<td>0.121</td>
<td>489</td>
<td>-1.021</td>
<td>p=1.61e-06</td>
</tr>
<tr>
<td>voiced.labiodentals</td>
<td>0.045</td>
<td>389</td>
<td>-1.092</td>
<td></td>
</tr>
<tr>
<td>voiceless.velars</td>
<td>0.011</td>
<td>54</td>
<td>-1.151</td>
<td></td>
</tr>
<tr>
<td>voiced.palatals</td>
<td>-0.066</td>
<td>102</td>
<td>-1.194</td>
<td></td>
</tr>
<tr>
<td>voiceless.palatals</td>
<td>-0.071</td>
<td>423</td>
<td>-1.208</td>
<td></td>
</tr>
<tr>
<td><strong>gender</strong></td>
<td></td>
<td></td>
<td></td>
<td>p=0.00092</td>
</tr>
<tr>
<td>FEM</td>
<td>0.084</td>
<td>657</td>
<td>-1.012</td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>-0.084</td>
<td>995</td>
<td>-1.181</td>
<td></td>
</tr>
<tr>
<td><strong>phone label</strong></td>
<td></td>
<td></td>
<td></td>
<td>p=0.0545</td>
</tr>
<tr>
<td>KIT</td>
<td>0.026</td>
<td>838</td>
<td>-2.10</td>
<td></td>
</tr>
<tr>
<td>FLEECE</td>
<td>-0.026</td>
<td>814</td>
<td>-1.14</td>
<td></td>
</tr>
<tr>
<td><strong>speaker (random)</strong></td>
<td>Std.Dev. 0.123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>word (random)</strong></td>
<td>Std.Dev. 0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance = - 505</td>
<td>df=16</td>
<td>Mean = - 1.114</td>
<td>AIC=1116,</td>
<td>Intercept = -1.097</td>
</tr>
</tbody>
</table>

Table 5.6: F1 results for KIT & FLEECE for the complex model (n = 1652). The $R^2$.fixed = 0.104, $R^2$.random =0.275

As a matter of convention in mixed effects modelling, $R^2$ values for random effects often tend to be higher than for fixed effects due to the likelihood of much larger variation triggered
primarily by the effects of individual words and speakers. Bearing this correlation in mind, while a range of values between 0.20 and 0.50 would be pretty standard for the $R^2$ fixed effect in results for L1 varieties, a value between 0.10 and 0.30 would be an expected range for L2 varieties (Hofmann, p.c). Hence, I compared the AIC as well as $R^2$ values among the six runs to determine the best performed.

The AIC for F1 for the following co-texts was 1574 and F2 = 1847; and the $R^2$ fixed were 0.11 and 0.10 respectively. For preceding co-texts, the AIC in F1 approximated to 1425 and 1689 in F2 with the $R^2$ fixed effects at 0.11 and 0.11. Next is the complex model (which includes the following and preceding co-texts in one run). The AIC for this model stood at 1116 for F1 and F2 =1615, with corresponding $R^2$ fixed effects at 0.10 and 0.10.

Evidently, the outcomes are consistent in lower AIC values which support the aptness of F1 measurement for determining the status of KIT and FLEECE, at least for this variety. In addition, the difference between the $R^2$ fixed values returned by the following, preceding and complex models were negligible, which means that any of the results would be essentially sufficient from the analysis.

With regard to significance hierarchy or the effect ranks, the results for the following co-texts and complex model were more consistent than some universal patterns for these classes of vowels. For example, both models returned the highest significant effect for phonological co-texts, then gender followed by phone label. And given the commonness of following co-text in the assessment of vowel status (see Hofmann 2015:273), I accepted its effect for phone label for KIT and FLEECE ($R^2= 0.10$ p= 0.00655) – which also corroborates the graphical indication of differentiation in Figure 5.2.

5.2.2 GOOSE and FOOT

Unlike the FOOT vowels, deciding on GOOSE tokens depends largely on the preceding consonantal environments. The GOOSE sets in the preceding yod /j/ and coronal environments are frontish, and realised as rounded central or front [u–y] (Johnston 1997b: 466, Labov, Ash & Boberg 2006b: 152, Mesthrie 2010: 10, Brato 2012:96). In the following environment, also, the retracting influence of /l/ has also been discussed as potentially misleading in the determination of GOOSE status. The fronting of GOOSE, therefore, results from articulatory pressures, which may not necessarily be indicative of a phonemic change in the system. The CELEX database, for example, shows that over 70% of /u/ vowels in English are preceded by
Figure 5.3: Scatter plot showing USE _GOOSE distribution in the F1&F2 (n = 635). The apparently outlying GOOSE tokens in the overlapping space with USE occur mostly in preceding and following lateral /l/ and liquid /r/ - which were excluded before statistical runs.

<table>
<thead>
<tr>
<th>following co-texts</th>
<th>FOOT</th>
<th>GOOSE</th>
<th>total</th>
<th>adjusted distribution following</th>
<th>FOOT</th>
<th>GOOSE</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>pause.pause.pause</td>
<td>0</td>
<td>31</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiced.apical.lateral</td>
<td>28</td>
<td>154</td>
<td>182</td>
<td>apicals</td>
<td>247</td>
<td>210</td>
<td>457</td>
</tr>
<tr>
<td>voiced.apical.stop</td>
<td>1</td>
<td>18</td>
<td>19</td>
<td>palatals</td>
<td>6</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>voiced.apical.stop</td>
<td>120</td>
<td>18</td>
<td>138</td>
<td>velars</td>
<td>67</td>
<td>23</td>
<td>90</td>
</tr>
<tr>
<td>voiced.labio-dental.frictive</td>
<td>0</td>
<td>23</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiced.labio-dental.liquid</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiced.palatal.frictive</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless.apical.stop</td>
<td>98</td>
<td>20</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless.labio-dental.frictive</td>
<td>0</td>
<td>23</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless.labio-dental.stop</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless.palatal.frictive</td>
<td>4</td>
<td>30</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless.velar.stop</td>
<td>67</td>
<td>23</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>320</td>
<td>357</td>
<td>667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: Initial cross-tabulation of all GOOSE & FOOT tokens with preceding phonological contexts before adjustment (n=677), and after adjustments (n=589).
/j/ or coronal sounds, both of which have preponderance for high F2 in the following vowel (Baayen, Dijkstra & Schreuder 1997, Mesthrie 2010, Hoffmann 2011). As a result, the combination of such GOOSE tokens with those occurring in relatively ‘non-fronting’ environments in the same model as a single class portends the odds of flawed results. Vowels in these environments are usually subsumed under the DEW and NEW classes (Brato 2012:96) or as ‘J-words’ for lexical items in the form of knew, use or dew (Mesthrie 2010:10).

The sub-categorisation thus allows for separate analysis of such tokens, and the avoidance of spurious findings for [u]. In this study (Figure 5.3), all GOOSE tokens preceded by /j/ were re-coded as USE during segmentation in PRAAT, and those in preceding and following liquid /l/ and lateral /l/ environments were later excluded from the regression runs.

As in the analysis of the high back vowels, different regression runs were constructed for following, preceding and both co-texts in F1 & F2. However, back vowels are predominantly differentiated in F2, thus more frequently assessed only in that dimension (Gorman and Johnson 2013:232, Hofmann 2015:273). While I ran regressions for F1 and F2 in order to accommodate possible system-specific behaviours that may be found in the data, only F2 results are explained from the models. Results from the runs were consistent with theoretical standpoints on back vowels – as the effect of phone label shows significance only in F2 across the co-texts. The reduction in the total tokens of GOOSE & FOOT (n=677 in Table 5.7) can be attributed to its lower frequency grid of occurrence and the exclusion of all GOOSE tokens before or after liquids or the lateral environments.

In following co-texts, the factors of place, manner and voice were merged as one variable (following co-texts) for regression analysis. Since there have not been theoretical accounts of consonants that might have certain conditioning effects on the high back vowels for EEng or generally NigE, the singularisation of the phonological environments was justifiable. As would be expected, the initial cross-tabulation of tokens according to following co-texts revealed gross inconsistencies.

Of the 13 levels, there were 6 with zero tokens and 3 with fewer than 5 tokens for either of the vowels. Unavoidably, these (and those in preceding and following pause environments) were removed, hence leaving just 3 factors within the variable. For the following co-texts, formant one and two were submitted to regression runs on 13 degrees of freedom with 589 tokens. The independent variables were as in previous analysis:
phonological co-texts, age group, education degree, style, number of L1, phone label and phone duration. And the individual speakers together word items were held as random.

In the second formant, phone duration showed a high significant effect ($R^2 = 0.21$ $p=0.000556$), phonological co-texts ($R^2 = 0.21$ $p=0.00509$) and phone label ($R^2 = 0.21$ $p=0.0144$). Education degree ($R^2 = 0.21$ $p=0.0749$), number of L1 ($R^2 = 0.21$ $p=0.0786$), age group ($R^2 = 0.21$ $p=0.189$), style ($R^2 = 0.21$ $p=0.378$) and gender ($R^2 = 0.21$ $p=0.438$) were insignificant. The overall behaviour in the following contexts thus suggests a differentiation of GOOSE & FOOT (Figure 5.4) mainly by linguistic factors (i.e., phone duration and phonological co-texts).

Figure 5.4: All speakers plot of FOOT & GOOSE tokens separated by preceding and following phonological co-texts (n=589). The vowels appear differentiated by the palatals and velars in F2, but apparently fused by apicals.
Table 5.8: Significant predictors for FOOT & GOOSE in F2 for preceding co-texts (n = 623) and complex model (n=589). Preceding co-texts: $R^2_{fixed} = 0.20$ $R^2_{random} = 0.51$. Complex model: $R^2_{fixed}=0.21$, $R^2_{random} = 0.52$.

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
<th>$R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>phone duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous +1</td>
<td>-0.079</td>
<td></td>
<td></td>
<td></td>
<td>$R^2 = 0.20$</td>
<td>p = 0.000556</td>
</tr>
<tr>
<td><strong>co-texts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>obstruct</td>
<td>0.334</td>
<td>109</td>
<td>-0.610</td>
<td></td>
<td>$R^2 = 0.20$</td>
<td>p=0.00509</td>
</tr>
<tr>
<td>liquids</td>
<td>-0.042</td>
<td>54</td>
<td>-1.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velars</td>
<td>-0.050</td>
<td>289</td>
<td>-1.228</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labials</td>
<td>-0.241</td>
<td>225</td>
<td>-1.249</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>phone label</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOOSE</td>
<td>0.092</td>
<td>357</td>
<td>-1.068</td>
<td></td>
<td>$R^2 = 0.20$</td>
<td>p=0.0511</td>
</tr>
<tr>
<td>FOOT</td>
<td>-0.092</td>
<td>320</td>
<td>-1.177</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>speaker (random)</strong></td>
<td>Std.Dev. 0.139</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>word (random)</strong></td>
<td>Std.Dev. 0.393</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance = -295</td>
<td>df=14</td>
<td>Mean = -1.119</td>
<td>AIC=668</td>
<td>Intercept = -1.009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
<th>$R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>phone duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous +1</td>
<td>-0.079</td>
<td></td>
<td></td>
<td></td>
<td>$R^2 = 0.21$</td>
<td>p = 0.000556</td>
</tr>
<tr>
<td><strong>co-texts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatals</td>
<td>0.339</td>
<td>42</td>
<td>-0.650</td>
<td></td>
<td>$R^2 = 0.21$</td>
<td>p=0.00509</td>
</tr>
<tr>
<td>velars</td>
<td>-0.162</td>
<td>90</td>
<td>-1.161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>apicals</td>
<td>-0.179</td>
<td>457</td>
<td>-1.179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>phone label</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOOSE</td>
<td>0.139</td>
<td>269</td>
<td>-1.094</td>
<td></td>
<td>$R^2 = 0.21$</td>
<td>p=0.0144</td>
</tr>
<tr>
<td>FOOT</td>
<td>-0.139</td>
<td>320</td>
<td>-1.177</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Speaker (random)</strong></td>
<td>Std. Dev. 0.131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>word (random)</strong></td>
<td>Std.Dev. 0.424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance = -252</td>
<td>df = 13</td>
<td>Mean = -1.139</td>
<td>AIC = 574</td>
<td>Intercept = -0.921</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The factors for the preceding co-texts comprised of 8 levels, i.e., comparatively fewer than the levels for the following co-texts. The removal of tokens in liquid and pause environments, and the subsequent merger of smaller factors into new related classes of phonological categories brought the levels down to a broad category of labials, obstruents and velars (n= 623). For F1, only gender \( (R^2 = 0.12 \ p=0.0108) \) and phonological co-texts \( (R^2 = 0.12 \ p=0.035) \) had significant effects according this model. The AIC was 291.86 with \( R^2 \) values for random and fixed factors at 0.97 and 0.306. The significance pattern for preceding co-texts in F2 was fairly similar to F2 in the following co-texts. Phone duration was highly significant \( (R^2 = 0.21 \ p=7.11e-06) \), and phonological co-texts \( (R^2 = 0.21 \ p=0.000424) \), then phone label \( (R^2 = 0.21 \ p=0.0511) \). The factors of age group \( (R^2 = 0.21 \ p=0.101) \), education degree \( (R^2 = 0.21 \ p=0.121) \), number of L1 \( (R^2 = 0.21 \ p=0.178) \), gender \( (R^2 = 0.21 \ p=0.309) \) and style \( (R^2 = 0.21 \ p=0.333) \) were not significant. The AIC was 629.51 with \( R^2 \) fixed and random values of 0.21 and 0.50. In comparison with the F1 model, the AIC is much higher but with a lower deviance at – 278.97 to -106.11.

Though the procedures were different, the results for the complex model (Table 5.8) were surprisingly similar to statistical outcomes for the following co-texts. Both models had exact number of tokens (n=589) and degrees of freedom after adjustments. The combination of factors in the following and preceding co-texts returned a staggering 154 levels after initial cross-tabulation with phone label. Tokens in liquid and pause environment, and levels with zero cells for either of the vowels were removed. The remaining factors were reduced to 3 levels – redistributed based on voice and place of articulation.

Again, the results of this model were quite parallel with results of the previous models. In F1, only gender \( (R^2 = 0.10 \ p=0.0199) \) and style \( (R^2 = 0.10 \ p=0.0409) \) were significant. Other variables were not significant. The AIC was 278.85 with deviance of 99.97. The \( R^2 \) fixed and random were 0.10 and 0.30. Results for F2 showed high significance effect for phone duration \( (R^2 = 0.21 \ p=0.000556) \), phonological co-texts \( (R^2 = 0.21 \ p=0.00509) \) and phone label \( (R^2 = 0.21 \ p=0.0144) \). The hierarchy of insignificance was very much similar to earlier results for following co-texts. The deviance was -251.57 and AIC = 574.41. The \( R^2 \) value for the fixed effects was \( (R^2 = 0.21) \) and random effects \( (R^2 =0.51) \).

Except in preceding co-texts where the effect of phone label was on the significance threshold \( (R^2 = 0.20 \ p=0.0511) \), results for other models suggest a differentiation of GOOSE from FOOT in F2. Considering the parallels between the statistical outcomes for the following
co-texts and complex model, only F2 results for the preceding co-texts and the complex model are shown in Table 5.8.

Though the runs seem very unanimous with reference to the number and order of significant variables, especially in F2, it is important to determine which among them best explained the data. The complex model in F2 comparatively ranked above other models with an AIC value of 574.41 and $R^2$ fixed of 0.21. Both models are much similar, almost by all parameters. More importantly, they uniformly returned high significant effects for phone label, which reiterates a distinguished status between GOOSE and FOOT in the F2.

With reference to my second research question, these outcomes suggest an interesting trend, especially in the light of attested behaviours in Educated Hausa and Yoruba English. A near RP-like differentiation between GOOSE and FOOT has been reported for Educated Hausa English, and an outright non-differentiation in Yoruba English (Banjo 1995, Simo Bobda 1995, Jibril 1995, Awonusi 2004). Though the Hausa L1 system operates a quantitative differentiation in high vowels, is unclear as to why they differ from their Southern counterparts (cf. Awonusi 2004 & Gut 2004). Unlike in Yoruba L1, the high back group in Hausa are phonemically contrastive, i.e., realised and perceived as different vowels (Malah & Rashid 2015:110), but qualitatively in Ebira (Adive 1989:25-6). However, the methodological divergence between my studies and these previous assessments makes the direct comparison of my findings with their problematic (cf. Hoffmann 2011).

In addition to quality differences between the vowels, both models yielded the highest effect for phone duration i.e., in preceding co-texts ($R^2 = 0.20 \ p=7.11e-06$), and in complex model ($R^2 = 0.21 \ p= 0.000556$) The high significant effects yielded for phone duration also suggest the tendency towards a shorter realisation of FOOT than GOOSE (in terms of ms), thus indicating a chiefly quantitative differentiation (see also Awonusi 1986:557).

5.2.3 Low Back Vowels: LOT, THOUGHT AND STRUT

The status of low back vowels (also often coded as COT and CAUGHT) is possibly the most interesting and widely studied indexical features of lectal variations in English varieties. The merger of both vowels or otherwise can be indicative of change either within or between systems, especially among those of established varieties in Western United States and other American varieties (Labov, Ash & Boberg 2006, Hall-Lew 2009, Nycz & Hall-Lew 2014). Long before the ANAE (Atlas of North American English) project which shows the typology
of low back vowels among some North American systems, Lusk (1976: 103-4) had observed the rapidly spreading non-distinctive status of LOT and THOUGHT lexical sets in Kansas City, especially for speakers born between 1940 and the early 1960s. Some analyses of San Franciscan variety, for instance, have found differing trajectories towards merger and differentiation (Moonwomon 1992, Labov, Ash and Boberg 2006, Hall-Lew 2013 & 2014), while studies on the Newfoundland variety of Canadian English have been unanimous on the palpable fusion of LOT & THOUGHT – dotted with regional inflections (Hollett 2006, Clark, 2004, 2010, Hofmann 2015). These and many other assessments of the low back vowels underscore its structural fluidness across geolinguistic or socio-lectal boundaries.

Similar to other systems where the low back vowels have been reported with phonological variations, the literature on NigE reflects a polarity between singularisation (Christopherson 1954, Jowitt 1991, Banjo 1995, Simo Bobda 1995, Eka & Udofot 1996, Udofot 2004 and Awonusi 2004) and quantitative differentiation (Jibril 1982, Eka 1985, Odumuh 1987, Eka 2000) of the vowels. While the claim of singularity seems fairly plausible, quantitative distinction between tense and lax vowels in English often finds explanation in the style of oral English teaching at the early years of schooling. Resulting from orthographical differences (as well as the phonemic length marks in long vowels), NigE learners are usually exposed to phonemic distinction in terms of duration rather than quality (Awonusi 1986:557). In what appear an inter-ethnic marker and structural similarity between NigE and the South Newfoundland English (SNLE), some Hausa English speakers conflate LOT with TRAP, rendering them as homophones (Jibril 1982, Clark 2010).

The STRUT class is often represented as a relatively short, half-open or slightly open, centralised back or central, unrounded vowel within the space of [ã] (Wells 1982:131). Depending on essentially demographic factors, STRUT has been reported to correlate strongly with either fronting or retraction in mainly British varieties (Wells 1982:292, 1994:121, Harrington et al. 2000:72, Torgersen & Kerswill 2004:40, Hughes et al. 2005:49). According to some of these studies, STRUT-fronting often triggers a partial merger with TRAP (Wells 1982:292), while its retraction may pull it into proximity with low back classes. As with LOT and THOUGHT, the non-distinction of STRUT from the low back cluster is also typical of NigE (Ekong 1987, Odumuh 1987, Simo Bobda 1995, Eka & Udofot 1996, Awonusi 2004 & Adetugbo 2004). Amidst widespread lack of distinction, however, the possibility of alternation between the short back LOT /ɔ/ and central TRAP /æ/ for STRUT /ʌ/ has been reported for certain speakers (Jibril 1982, Jowitt 1991 & Eka 2000). This notion, however, seems to
find a logical reference in Daniel Jones’s advice to foreign learners— which suggests a substitution with TRAP [a], ‘if all efforts to obtain the precise sound STRUT /ʌ/ fail’ (Collins & Mees 1999:23, Fabricius 2007:296). As far as NigE stands described, the dearth of central vowels in the lower ‘lects’ is resultant in the absence of STRUT and lettER, given their equivalent absence in the primary (L1) systems. Given the claims of non-differentiation between LOT, THOUGHT and STRUT in NigE, it becomes necessary to evaluate their status, especially in the context of current data.

Low back vowels are more robustly separated in F2 (Gorman & Johnson 2013:232, Hofmann 2015:244), therefore assessed only in that dimension. So, regressions were run, initially with the adjusted distribution of LOT and THOUGHT tokens in following, preceding and both co-texts. Studies on some established varieties support the sway of phonological co-texts on the low back sets as most remarkable in the following co-texts. For example, Labov, Ash & Boberg (2006:59) observe that, more than other consonantal co-texts, the low back merger is most commonly favoured in the following nasal sounds— due to its backing effects (cf. Thomas 2011:101, Hofmann 2015:216). The picture is not as clear-cut in the preceding environment, despite the raising effects of the flanking glides, liquids and oral labials in F2. Generally, all tokens in glide environments (except USE tokens which I retained as a separate sub-set of GOOSE), approximants and liquids were excluded in the regression analysis. Tokens in preceding oral labials were however retained due to the lack of consensus as to the degree of effects on F2 (Table 5.9).

A total of 1659 tokens (LOT = 850, THOUGHT = 809) were returned after initial cross-tabulation of the following co-texts with phone label. Though the distribution appears fairly even, a look into the table revealed some internal discrepancies. Of the total 40 cells for both sets, 6 had zero tokens and about 13 cells with less than 10 tokens. Thus, factors with relatively smaller amount of tokens were combined with similar co-texts while ‘odd’ factors that could not be left in the model were removed. The cleaning and re-assigning of tokens in the following co-texts yielded four levels viz. apical, labio-dental, palatals and velars comprising 1444 tokens.

According to this model, the effects of phone duration ($R^2 =0.05 \ p=7.4e-09$) and style ($R^2 =0.05 \ p=0.0121$) were significant. The $R^2$ fixed and the $R^2$ random were 0.05 and 0.27 respectively, while the AIC was at 330. In the preceding model, the effect of phonological co-texts was significant ($R^2 =0.06 \ p=0.00175$) as well as style ($R^2 =0.06 \ p=0.0213$) and phone
duration ($R^2 = 0.06$ p = 7.31e-08). The AIC was 368 with $R^2$ fixed and $R^2$ random values of 0.06 and 0.22.

Figure 5.5: F1 & F2 contour plot of LOT and THOUGHT in the following and preceding contexts. The vowels have show equal distribution and complete fusion in both dimensions (n = 1651)
### Preceding Co-texts

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coefficient</th>
<th>Tokens</th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phone duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous +1</td>
<td>-0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CO-texts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apicals</td>
<td>0.050</td>
<td>213</td>
<td>-0.976</td>
<td></td>
</tr>
<tr>
<td>Obstruents</td>
<td>-0.030</td>
<td>262</td>
<td>-0.971</td>
<td></td>
</tr>
<tr>
<td>Velar</td>
<td>-0.005</td>
<td>728</td>
<td>-0.972</td>
<td></td>
</tr>
<tr>
<td>Labials</td>
<td>-0.075</td>
<td>241</td>
<td>-1.101</td>
<td></td>
</tr>
<tr>
<td><strong>Style</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>0.031</td>
<td>313</td>
<td>-0.989</td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>-0.031</td>
<td>1131</td>
<td>-0.995</td>
<td></td>
</tr>
<tr>
<td><strong>Phone label</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thought</td>
<td>0.007</td>
<td>737</td>
<td>-0.974</td>
<td></td>
</tr>
<tr>
<td>Lot</td>
<td>-0.007</td>
<td>707</td>
<td>-1.015</td>
<td></td>
</tr>
</tbody>
</table>

| Speaker (random) | Std. Dev. 0.083 |     |     |         |
| Word (random)    | Std. Dev. 0.115 |     |     |         |
| Deviance = -137  | df = 14       | Mean = -0.993 | AIC = 369 | Intercept = -0.994 |

### Complex Model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coefficient</th>
<th>Tokens</th>
<th>Mean</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phone duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous +1</td>
<td>-0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Style</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>0.028</td>
<td>418</td>
<td>-0.991</td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>-0.028</td>
<td>1233</td>
<td>-1.000</td>
<td></td>
</tr>
<tr>
<td><strong>CO-texts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatals</td>
<td>0.055</td>
<td>339</td>
<td>-0.883</td>
<td></td>
</tr>
<tr>
<td>Velars</td>
<td>0.010</td>
<td>171</td>
<td>-0.983</td>
<td></td>
</tr>
<tr>
<td>Apicals</td>
<td>-0.021</td>
<td>921</td>
<td>-1.032</td>
<td></td>
</tr>
<tr>
<td>Labio-dentals</td>
<td>-0.043</td>
<td>220</td>
<td>-1.042</td>
<td></td>
</tr>
<tr>
<td><strong>Phone label</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOT</td>
<td>0.002</td>
<td>849</td>
<td>-1.008</td>
<td></td>
</tr>
<tr>
<td>THOUGHT</td>
<td>-0.002</td>
<td>802</td>
<td>-0.986</td>
<td></td>
</tr>
</tbody>
</table>

| Speaker (random) | Std. Dev. 0.082 |     |     |         |
| Word (random)    | Std. Dev. 0.111 |     |     |         |
| Deviance = -117  | df = 14       | Mean = -0.998 | AIC = 330 | Intercept = -0.983 |

Table 5.9: LOT & THOUGHT in F2 for preceding co-texts (n = 1444), and complex model (n=1651). Preceding co-texts: $R^2$.fixed =0.58 $R^2$.random =0.22. Complex model: $R^2$.fixed=0.05. $R^2$.random = 0.22.
The combination of both phonological environments (complex model) initially resulted in a staggering variable of 104 factors. These, however, were subsequently reduced to the same number of consonantal parameters as in the previous models at a total of 1651 tokens. The outcomes yielded by this model was very much similar to those of the following co-texts’. Again, only phone duration \( R^2 = 0.05 \ p=7.4e-09 \) and style \( R^2 = 0.05 \ p= 0.0121 \) were significant, while the AIC and deviance values were very similar to those of the following co-texts’. Consequently, the details of both models were straightforwardly compared with those of the preceding co-texts in order to determine which the best explained the variances. The AIC value for the following co-texts mode was 330 with \( R^2 \) fixed value of 0.05, while for the preceding co-texts, the AIC was 368 with \( R^2 \) fixed value of 0.06. Strikingly, both models appeared to have behaved much alike in terms of overall statistical outputs.

Considering the importance of phonological environments in the assessment of low back vowels (Labov, Ash & Boberg 2006, Thomas 2011), the results in the preceding co-texts (which yielded a high significant effect for phonological co-texts) can be adjudged to have excelled in assessing the data. The models regardless, the differentiating effect of phone duration was clear from the results. Phone label was however not significant in any of the models (following co-texts: \( R^2 =0.05 \ p=0.826 \), preceding co-texts \( R^2 =0.06 \ p=0.591 \) & the complex model \( R^2 =0.05 \ p = 0.826 \), thus clearly confirming an identical status for LOT & THOUGHT. This finding, in effect, appeared to have harmonised the polarity of notions with regard to the realisation of these vowels in NigE – supporting the quantitative differentiation of the vowels (Jibril 1982, Eka 1985, Odumuh 1987 & Eka 2000) and the qualitative non-differentiation (Jowitt 1991, Eka & Udofot 1996, Udofot 2004 & Awonusi 2004).

5.2.4 STRUT and Low Back Vowels

The contour plot (Figure 5.5) and regression runs for LOT and THOUGHT has shown the vowels as undifferentiated, hence assigned as one phoneme. So, all tokens of LOT and THOUGHT were re-coded as a new class of THOUGHT, after which it was compared with STRUT. The objective was to determine whether the entire class of low back vowels exist as a single homophonous
cluster, or shows the route of gradience in F2 between them (Collins & Mees’s 1999:23, Fabricius 2007:96). The two sets consist of 3403 tokens (thought = 1659). The combination of the two vowel groups into thought created no artificial imbalances between the two new sets, as strut (n=1744) still had more tokens than thought (n=1659) after initial cross-tabulation with the following co-texts. The same procedure followed in the analysis of lot & thought was reapeeted in adjusting the tables for strut. The 20 factors representing: place, manner and voice for the following co-texts were trimmed down to 4 – based on much broader classification according to place (apicals, labio-dentals, palatals and velars). The predictive elements of previous models were retained, with F2 as the only outcome variable (Gorman & Johnson 2013:232, Hofmann 2015:244).

For the following co-texts model, the order of significance was led by phone duration ($R^2 =0.07 p=6.54e-09$), then phonological co-texts ($R^2 =0.07 p=2.34e-05$), phone label ($R^2 =0.07 p= 0.000748$) and style ($R^2 =0.07 p=0.00151$). The AIC was 2135 with $R^2$ fixed and random values of 0.07 and 0.19. Other predictors in this model yielded no significance. The preceding co-texts had same factors as in the following co-texts, viz. labials, apicals,
obstruents and velars. The dependent and independent variable were as in previous runs, while the total number of tokens was 2970: STRUT (n=1526) and THOUGHT (n=1444). Again, phone duration had the highest significant effect ($R^2 =0.07 \ p=3.12e-07$), before phonological co-texts ($R^2 =0.07 \ p=0.000546$), style ($R^2 =0.07 \ p=0.00247$) and phone label ($R^2 =0.07 \ p=0.00891$). As in the results for the following environments, the effects of gender, age-group, education degree and number of L1 were insignificant. The deviance and AIC values for this model approximated to –994 and 2084. The $R^2$ values for fixed and random effects also were 0.07 and 0.20 respectively.

Results for the complex model returned significant values for the same number of variables included in the analysis of the following and preceding co-texts. Phone duration was most significant ($R^2 =0.26 \ p=7.03e-09$), followed by phonological co-texts ($R^2 =0.26 \ p=7.09e-05$), phone label ($R^2 =0.26 \ p=0.000531$) and style ($R^2 =0.26 \ p=0.00174$). The AIC and $R^2$ values were quite similar to those of the following co-texts.

Results from the different models are interestingly alike in terms of significant factors. For example, the AIC values were as follows: following co-texts = 2135, preceding co-texts = 2084, and complex model = 2137. Given the negligible discrepancies between these models and the correspondence between their significant predictors, any of the results would be sufficient in terms of how accurately the variances had been accounted for.

Table 5.10 shows the results for STRUT and THOUGHT based on the complex model. Unlike for LOT and THOUGHT, it supports a contrastive realisation between STRUT and THOUGHT in the data. It is however unclear if such difference as revealed by statistical analysis in this study is merely superficial, or arising from factors to which the regression models are blind. Likelihoods also are that such effects, particularly of phone label, originated from the backing and lowering of STRUT at the same time – in the sense that speakers tend to alternate between TRAP and STRUT (cf. Jibril 1982, Jowitt 1991 & Eka 2000). There are, however, some striking parallels in the analysis of LOT and THOUGHT on one hand, and STRUT and THOUGHT on the other. Differentiation between the two pairs is most pronounced in terms of duration. Phonological co-texts and style also had significant effects for the entire models, therefore suggesting a strong distinctive influence of linguistic factors between the vowels in the system. Figure 5.6 shows the pattern of STRUT/THOUGHT realisations in formal and informal styles of speech. And as would be expected, the variation between the speech styles is explainable (Labov 1990:224).
5.2.5 Low Central Vowels

The TRAP /æ/ and BATH/aː/ classes were primitively achieved as one vowel [a] in Middle English (ME). As a result of what is now known as pre-fricative lengthening, [a] which had earlier been raised to [æ] later became lengthened in the following [f], [θ] or [s] environments. Similarly, the lengthening process was extended to the Middle English /ɒ/ which sounded as /ɒː/ - also in following fricative contexts (Wells 1982:204). Thus, the two classes were, in effect, realised as ‘contextual variant of one phoneme’ (Roca & Johnson 1999:683). The quantitative difference between the two vowels was marked to have also gradually assumed quality distinction amid lexical inconsistencies by the close of 17th century (Wells 1982: 204-233, Beal 1999:105). Subsequently, the TRAP/BATH classes have undergone phonemic mutations among varieties of English and remain heterogeneous till today. For example, while the TRAP/BATH lexical sets are characterised by a single vowel in Scottish Standard English (ScSE) and in Northern English, TRAP commenced a development toward a split in Southern English between the mid-18th and the early 20th centuries – such that the vowel became variably backed and/or lengthened in following nasal or voiceless obstruent environments.
(Wells 1982:205, Piercy 2011: 155, Nycz & Hall-Lew 2014: 10-11). In the process of change, however, the TRAP/BATH split never went to completion, but seemed to have stopped before spreading its realisation through the English lexicon (Wells 1982:233). Also, the phonological context in which the two vowels are separated is highly ambiguous and random. The fall-out of these developments in today’s English is often reflected in pairs like gaff & crass – staff & brass or trample – sample where the first of each set is represented as TRAP and the second as BATH, thus making the disparity between the two vowels specially knotty, even among the native speakers (Nycz & Hall–Lew 2014:11). Speakers often have phonetic difficulties as to which of the two sounds to use in certain co-texts, as well as the right words in which to employ either of the sounds, resulting in proneness to oscillation between p[æ]st and p[ɑː]st or dem[æ]nd and dem[ɑː]nd, for example (see Lloyd 1895:53).

In what appears the most ambitious of earliest characterisation of West and East African English vowels, Schmied (1991a and b) reports an outright non-distinction for TRAP and BATH. His study was later corroborated in a follow-up comparison of formant values by Mutonya (2008:242-5), and the analysis of Black Kenyan English (Hoffman 2011:164). However, Lass’s (2004:106) description of TRAP in South African accent finds close equivalences in Australian (AusE) and New Zealand (NZE) English – both which have [æ] or [ɛ] for TRAP, and the low front [ɑː] for BATH; but further defines the centralisation of BATH [ǣ:] rather than the back [ɑː] for BATH in SAE. A broader review of previous reports on the TRAP/BATH continuum in NigE can be found in Section (2.5.1). While these studies are undivided on a complete non-distinction of the pairs in terms of quality, they disagree in the scale of variations and trajectory of differentiation within the system. Consistent with findings on the East and West African varieties (see Schmied 1991, Simo Bobda 1995, Mutonya 2008 & Hoffman 2011), a single low vowel in TRAP and BATH is attested in Christopherson (1954), Bamgbose (1995), Eka & Udoofot (1996) and Adetugbo (2004); while quantitative differentiation of the two sets is hinted in Eka (1985) & Jowitt (1991). In terms of internal variations for TRAP, Jibril (1982) and Awonusi (2004) confirm the possibility of [æ], [ɛ] & [eː], and a centralized [ɑː] for BATH as alternative for some speakers (cf. Lass 2004: 106). Given the nebulousness of TRAP and BATH; and the paucity of empirical portrayal of the NigE system, a further acoustic corroboration is in order (cf. Mutonya 2008 & Hoffman 2011 for East African varieties).

Similar to other sets of central or reduced vowels, the status of the schwa /ə/ (of the letter set) has also been variedly defined for NigE systems. The letter vowel has been
attested as homophonous with the low central classes i.e., TRAP & BATH, and occasionally with LOT and DRESS (Bamgbose 1995, Eka & Udofot 1996, Adetugbo 2004, Awonusi 2004). Since these studies were mainly based on auditory judgements, it is conceivable to finger what the underlying assumptions were. First, due to the influence of L1 (mostly implicated in the analysis of NigE phonology), the possibility of a non-stressed, weak /ə/ is rarely imagined (Jowitt 2000, Gut 2002). Second, given the fluidity of orthographical representations of the schwa /ə/ in words, the letter vowel tends to pose some pronunciation conflicts, particularly for speakers whose oral awareness of English words is predominantly spelling-driven (Awonusi 1986:556). However, in a recent auditory evaluation of educated NigE speakers, the letter in unstressed syllables is perceived as slowly losing its perceptual force to the adjacent stressed classes in words (Jowitt 2015:42). While a high possibility of appropriate weakening (i.e., schwa realisation) is attested in Educated Hausa English, the pattern is generally rare for the Southern varieties (Jibril 1986, Udofot 1990, Jowitt 1991). This notion, most likely, inspired a following perceptual study by Akinjobi (2006) in which she compared the percentage of schwa weakening by 100 Educated Yoruba English speakers with that of a native British speaker. More details of her findings have been discussed in Section 2.5.1. A total of 92% of her respondents demonstrate overly similar behaviours as in the stressed syllables. This is supported by the high similarity between the duration mean values of schwa-syllables and neighbouring constituents: 141.9ms and 144.4ms respectively (Akinjobi 2006:13).

With regard to current data, the foregoing thus present some interesting prospects. On the one hand is the affordance of substrate possibilities in the achievement of schwa; and on the other is its presence in the Hausa or the Northern varieties. In his comment on the formation of vowels within a single syllable in Ebira L1, for instance, Adive (1989: 17) defines the second item in the germinate sequences as a schwa (ə), thus reinforcing what Carnochan (1970:224) describes as ‘structural schwasyllable prosody’ in some Southern Nigerian languages. The Ebira people, more importantly, have a history of shared linguistic and cultural affinities with the North and the South of Nigeria. Wedged between these regions, the earliest forms of English or formal education came from the South. The people, however, later developed a strong cultural and linguistic fancy for the North (Section 2.2). Thus, the status of letter, as shown in the current analysis is instructive as to the direction the system is most likely accommodating.
5.2.5.1 Analysis of TRAP, BATH and letter

First, separate regression for TRAP and BATH were fitted for following, preceding and complex models. Unlike for back vowels, the literature is largely undecided on the most robust dimension for assessing the status of TRAP and BATH. Also, the visualisation of these vowels as Figure 5.7 suggests neither F1 nor F2 as potentially more suitable for analysis. Concerning consonantal influence, however, the effect of following co-texts has been attested (Wells 1982:205, Piercy, 2011:155, Nycz & Hall-Lew 2014:10-11). And it is against the theoretical backdrop that only the following co-texts were considered in F1 & F2 for TRAP and BATH.

Figure 5.7: F1/F2 plot showing similar pattern of distribution for TRAP & BATH in both dimensions (n =1998)

As in previous analysis, tokens were re-assigned only according to place of articulation in the following co-texts: apicals, labio-dentals, palatals and velars. The total number of vowels came down to 1976 after re-adjustment to the raw distribution. For F1, only phone duration ($R^2 = 0.10$ $p=2.08e-09$) and style ($R^2 = 0.10$ $p=3.29e-07$) had significant effects. The values for phone label ($R^2 = 0.10$ $p=0.375$) and other variables were however insignificant. The AIC was 2653. The $R^2$ fixed and random were 0.082 and 0.191. In F2,
however, the analysis showed the highest effect for phone label \( R^2 = 0.20 \ p = 5.33 \times 10^{-6} \), then phonological co-texts \( R^2 = 0.20 \ p = 0.00078 \), style \( R^2 = 0.20 \ p = 0.000927 \) and gender \( R^2 = 0.20 \ p = 0.00155 \). Conversely, the AIC for the F2 run yielded a negative value of – 146. The \( R^2 \) fixed and \( R^2 \) random were 0.20 and 0.37. While the F2 results had lower AIC values, F1 seemed to have performed better based on the deviance \( R^2 \) total (fixed and random factors). Results for the complex model were much similar to outcomes for the following phonological co-texts in both dimensions.

Though either of F1 or F2 outcomes would find correspondence in previous assessment of TRAP and BATH in NigE, the illustration in Figure 5.7 appears to favour a single system. This was however complicated by the discrepancy between the statistical results for F1 and F2. While the regression indicated an identical realisation in F1, a differentiation was supported by the F2 results. Given the blindness of the regression runs to correlations between the two formants – and in fact, its inability to simultaneously account for two dependent outcomes in a single model, neither of these results could be taken seriously.

Consequently, I decided on a MANOVA – which, fully, could compare the vowels’ F1 and F2 in a single fit. But unlike mixed effects regression, MANOVA is less forgiving – especially when its assumptions are not fairly met (Section 4.2.1). The most important of these assumptions, perhaps, is the independence of observations – which in essence requires the assignment of each token to one speaker in the data. For the current data, separate mean values of TRAP and BATH tokens were calculated for each speaker. This procedure reduced the total number of observations to \( 58 = \frac{58}{2} \), such that there were 29 tokens for each class. As explained in Section 4.2.1, a MANOVA is most suitable with relatively larger amount of data. The severity of this assumption however depends on the goal of analysis, or alternatively, the number of levels in the independent variables under assessment (Hofmann, p.c). The main predictors tested against the dependent variables (F1 & F2) were phone label, gender and age group (each of whose level was binary), thus precluding the possibility of Type 1 error in a MANOVA results for the data. The correlation assessment between the F1 and F2 fell within the healthy range of – 0.4 and – 0.9 (Pearson’s \( r = -0.38 \)) in the case of negative correlation (Mayers 2013:323, Hofmann 2015:217).

In general, the extent to which MANOVA assumptions are met can be verified via an omnibus tool known as global test (Peña and Slate 2003). The global test is usually recommended before linear runs – in order to determine: the skewness of samples, kurtosis of the error distribution, the use of link function, and the heteroscedasticity of variances between
the samples (Peña and Slate 2003:11, Hofmann 2015:217). Table 5.12 illustrates the outcomes of global tests on four degrees of freedom in F1 & F2 for TRAP and BATH. The tests give a clean bill of health regarding all assumptions, with the exception of the violation of kurtosis for BATH in F1. The Global Stat however indicates a general fulfilment of this requirements (value = 8.939e+00 pvalue = 0.06265).

<table>
<thead>
<tr>
<th>TRAP_F1</th>
<th>Value</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Stat</td>
<td>1.686e+00</td>
<td>0.7933</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.279e-04</td>
<td>0.9880</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.044e-03</td>
<td>0.9560</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Link Function</td>
<td>7.526e-13</td>
<td>1.0000</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>1.682e+00</td>
<td>0.1946</td>
<td>Assumptions acceptable.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TRAP_F2</th>
<th>Value</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Stat</td>
<td>3.996e-01</td>
<td>0.9825</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Skewness</td>
<td>3.050e-01</td>
<td>0.5807</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>8.711e-02</td>
<td>0.7679</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Link Function</td>
<td>2.229e-16</td>
<td>1.0000</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>7.409e-03</td>
<td>0.9314</td>
<td>Assumptions acceptable.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BATH_F1</th>
<th>Value</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Stat</td>
<td>8.939e+00</td>
<td>0.06265</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.213e+00</td>
<td>0.13687</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.426e+00</td>
<td>0.03540</td>
<td>NOT satisfied!</td>
</tr>
<tr>
<td>Link Function</td>
<td>-1.785e-13</td>
<td>1.00000</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>2.300e+00</td>
<td>0.12937</td>
<td>Assumptions acceptable.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BATH_F2</th>
<th>Value</th>
<th>p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Stat</td>
<td>1.434e+00</td>
<td>0.8382</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Skewness</td>
<td>3.686e-01</td>
<td>0.5438</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.232e-01</td>
<td>0.7256</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Link Function</td>
<td>5.025e-16</td>
<td>1.0000</td>
<td>Assumptions acceptable.</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>9.424e-01</td>
<td>0.3317</td>
<td>Assumptions acceptable.</td>
</tr>
</tbody>
</table>

Table 5.11: Results of linear model assumptions for a MANOVA regarding the distribution of gender and mean formant values of TRAP and BATH on four degrees of freedom (n = 29). The assumptions are met for all the tests, except for Kurtosis in BATH_F1. Level of significance is set to 0.05

A MANOVA based on Wilks' lambda was run for variance in the first and the second formants, including phone label, age and gender as predictors. The model returned non-significant effects for phone label $F(1,54) = 2.24$, $p = 0.12$ (Wilks' $\lambda = 0.92$), and age $F(1,54) = 2.11$, $p = 0.14$ (Wilks' $\lambda = 0.94$). The effect of gender was however significant: $F(1,54) = 10.30$, $p = 0.0001$ (Wilks' $\lambda = 0.72$). With these predictors, the univariate differences
between the phonemes based on post-hoc ANOVA showed insignificant effect for phone label in F1: $F(1,54) = 2.22, p = 0.14$, gender: $F(1, 54) = 2.023, p = 0.16$, and age: $F(1,54) = 3.61, p = 0.0629$. The effect was also not significant in F2 for phone label: $F(1,54) = 2.36, p = 0.13$ and age: $F(1,54) = 1.15, p = 0.29$. As in the MANOVA employed, the effect for gender was again highly significant: $F(1,54) = 18.36, p = 7.58e-05$.

As earlier explained, a follow-up MANOVA for TRAP and BATH became necessary – owing to the major incapacity of the linear regression run to take both the first and second formants into account in one analysis. Barring a laundry of statistical advantages amply exploitable in the regression analysis, this inability was compensated for by MANOVA, and the post-hoc ANOVA for univariate effects of the factors in F1 and F2. The MANOVA results indicate a non-differentiation of TRAP and BATH, thus supportive of earlier findings in Bamgbose (1995), Eka & Udofot (1996), Adetugbo (2004), Udofot (2004). While the claim of quantitative distinction (which also finds support in the effects of phone duration), the two vowels are achieved with a much similar quality.

Based on these results, the two sets could hence be assigned as one phoneme in the system – which means that their non-differentiation from letter might be further assessed with either of TRAP or BATH. A standard procedure would be to recode the two sets into one vowel class, and be compared with letter. The subsumation however widened the disproportions between the BATH (n=1987) and letter (n= 401) tokens in a scale that could warrant erroneous results. Subsequently, I made a decision to choose either of TRAP (n= 1209) or BATH (n= 778) with a lesser amount of tokens. Since both vowels have an identical phonemic status in the system, BATH was chosen, and compared with letter. Based on the confirmation of the F1 results for TRAP & BATH by the post-hoc ANOVA, I decided on the suitability of F1 for further analysis of low central vowels (c.f Nycz & Hall-Lew 2014:11, Hofmann 2015:254). In the following co-texts, phone label was most significant ($R^2= 0.42 p=2.14e-12$), followed by style ($R^2= 0.42 p=6.54e-10$) and phone duration ($R^2= 0.42 p=6.34e-05$).
According to this model, phonological co-texts, surprisingly, was least significant of the variables \( R^2 = 0.42 \) \( p=0.777 \) with an AIC value of 1273. Similarly, the significance order was led by phone label in the preceding co-texts \( R^2 = 0.39 \) \( p=1.12\cdot10^{-9} \), then phone duration \( R^2 = 0.39 \) \( p=2.2\cdot10^{-9} \), style \( R^2 = 0.39 \) \( p=6.85\cdot10^{-8} \). The phonological environment was however significant according to this run \( R^2 = 0.39 \) \( p=0.0325 \). The AIC was also similar to the other run at 1276.35.

5.2.5.2 The Status of lettER and BATH

For lettER and BATH, all the regression analysis returned high significant effects for phone label in F1 (and in F2 as well), thus indicating a differentiation between the two vowels. However, a good number of lettER tokens were lost after adjustment to the initial cross-tabulation against the following co-texts, which resulted in further skewness between the two classes: BATH \( n=767 \) and lettER \( n=267 \). Although the deferential posed little or no threat
to the analysis – given the robustness of mixed effects modelling against inter-factor imbalances (Johnson 2009, Nycz & Hall-Lew 2014: 2), its statistical outcome would call for cautious weighing. Relatively, the distribution between the two classes in the preceding environments was however relatively balanced, i.e., BATH (n=652) and letter (n=370). The similarity of AIC and deviance values for BATH and letter (1273, –593.79) and (1276.35, –594.93) respectively in the models was fairly corroborated by the number and significance hierarchy in their results. More importantly, both models showed phone label as most significant among other predictors, thus making their outcomes interchangeably reliable.

The possibility of schwa weakening which, in effect, results in the pattern shown for this data has earlier been reported for Hausa English (Jibril 1986, Udofot 1990, Jowitt 1991). This notion has further reinforcement in Jowitt (2014:14) who hints on the gradual weakening of letter in unstressed syllables, thus supporting the presence of schwa /ə/ in the system. In an earlier quantitative comparison between the schwa and the tense [a] in Educated Yoruba English variety, Akinjobi (2006:13) reports a systematic non-differentiation between schwa-syllabic vowels and their adjacent constituents. The fact that her findings derived mainly from syllabic duration rather vowels’ formants would make comparison with then current findings problematic.

Based on auditory evaluations, however, the preponderance of schwa realisation as the centralised [ä:] has been attested in Odumuh (1985) and Eka (1985). In a later study, while Eka (2000) also enlists the possibility of schwa /ə/ among the variants of [ä:] for some speakers, Awonusi (2004) and Adetugbo (2004) report the low back vowels /ɔ, ɒ/ as more frequent variants of /ə/.
Figure 5.9 (a&b): Plots of letter and bath vowels in the following co-texts (n = 1030), indicating some overlap and a differentiation in F1: by phone label ($R^2 = 0.31$, $p=2.14e-12$), and style ($R^2 = 0.31$, $p=6.54e-10$); being predictors with the most significant effects in the regression runs. The total number of speakers = 29
5.2.6 NURSE

The NURSE vowel, depending on certain factors, has perhaps, the fluidest realisation in NigE and some West African varieties (Section 2.5.4). In two major surveys by Schmied (1991a) and Simo Bobda (2000), both studies agree on the amorphous status of NURSE – mostly conditioned by idiolectal, stylistic, orthographical and geolinguistic factors. According to Simo Bobda’s finding, the phoneme remains ‘the most distinguishing parameter in the regional, national and even ethnic identification of a speaker’ (2000:41). From the data available to him, he defines a range of realisation patterns chiefly marked by national and ethnic boundaries. Table 2.11 presents the summary of these phenomena and their internal nuances particularly for Nigerian speakers, while Table 2.9 highlights the possibilities of NURSE as already reported for NigE (Eka 1985, 2004, Odumuh 1987, Jowitt 1991, Adetugbo 2004). In Yoruba English for instance, Simo Bobda insists on the prospect of /ɔː/ for NURSE words with orthographic <or, ur and our>, and fluctuation between /a/ and /e/ in items with <ear, ir and er>. While the trend among Igbo speakers is generally similar to these tendencies, Hausa speakers demonstrate a general restructuring of /ɜː/ to /a/, regardless of orthographic markers (Simo Bobda 2000:42). The literature has therefore been fairly consistent in the backing of /ɜː/ to /a/, lowering to /a/ or fronting to /e/, hence suggesting a general propensity towards phonemic overlap between the traditional NURSE classes with THOUGHT, DRESS and BATH vowels.

Hence, my investigation was based on these prior accounts, so as to understand the measure to which NURSE approximates in the orthographical contexts, and the effects of other variables within the system. Section 5.2.4 discusses the general estimation of the external effects as well as their degree of correlation with previous findings. The commonness of allophonic overlap of NURSE with THOUGHT, DRESS and BATH has been attested. And in line with my research question, the pairs of NURSE & BATH, NURSE & DRESS and NURSE & THOUGHT were measured so as to determine their extent of coalescence or differentiation between each of them. Since lexical orthography has been consistently marked as major delineating parameter, I re-grouped all tokens of NURSE according to these spellings, e.g., <ear, ir and er> as NURSE_BATH and NURSE_DRESS; and <or, ur and our> as NURSE_THOUGHT. This splitting was also supported by the distribution of individual tokens of NURSE in the data. Consequently, an additional column with all NURSE tokens in their fresh grouping was created alongside BATH, THOUGHT and DRESS. The choice of BATH and THOUGHT for the low central and back region was based on the non-differentiation of the
vowels from their lax pairs – which resulted in their classification as single phonemes in the data.

Based on the foregoing, phone label, was replaced with ‘realisation_NURSE’. The new variable had 6 factors or levels comprising: BATH, THOUGHT, DRESS, NURSE_BATH, NURSE_THOUGHT and NURSE_DRESS. Given the movement of tokens within the vowel envelope, the variance between BATH & NURSE_BATH and between DRESS & NURSE_DRESS was measured in F1, but in F2 for THOUGHT and NURSE_THOUGHT. Accordingly, age-group, gender, education degree, realisation_NURSE, phonological contexts, phone duration and style were entered as fixed predictors with the effects of individual speaker and word as random. The distribution of these tokens after cross-tabulation and adjustments is shown in Table 5.13. For NURSE_BATH, the effect of realisation_NURSE was most significant in \( R^2 = 0.32 \ (p=8.32e-11) \), and style \( R^2 = 0.32 \ (p=5.51e-07) \), then phone duration \( R^2 = 0.32 \ (p=0.00113) \) on 12 degrees of freedom. The profound lowering of NURSE_BATH in F1 as reinforced by the strong significant value for realisation_NURSE was also supported by a large mean difference between the two classes: BATH (1.574) and NURSE_BATH (0.584) at 0.99.

For DRESS and NURSE_DRESS, the effect of realisation_NURSE was again most significant \( R^2 = 0.30 \ (p=3.83e-33) \). Additionally, the factors of style \( R^2 = 0.30 \ (p=4.65e-09) \), phone duration \( R^2 = 0.30 \ (p=1.18e-08) \) and phonological co-texts \( R^2 = 0.30 \ (p=0.0517) \) also yielded significant values on 12 degrees of freedom. The evaluation of THOUGHT and NURSE_THOUGHT basically followed a similar procedure, except that the dependent variable was assessed in F2. The effects of style \( R^2 = 0.15 \ (p=8.87e-06) \), phone duration \( R^2 = 0.15 \ (p=1.68e-05) \) and phonological co-texts \( R^2 = 0.15 \ (p=0.00186) \) were highly significant.

<table>
<thead>
<tr>
<th>co-texts</th>
<th>BATH</th>
<th>DRESS</th>
<th>THOUGHT</th>
<th>NURSE_BATH</th>
<th>NURSE_THOUGHT</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>apicals</td>
<td>63</td>
<td>998</td>
<td>408</td>
<td>173</td>
<td>50</td>
<td>1692</td>
</tr>
<tr>
<td>labio-dentals</td>
<td>118</td>
<td>468</td>
<td>50</td>
<td>70</td>
<td>34</td>
<td>740</td>
</tr>
<tr>
<td>palatals</td>
<td>176</td>
<td>357</td>
<td>268</td>
<td>135</td>
<td>156</td>
<td>1092</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>357</strong></td>
<td><strong>1823</strong></td>
<td><strong>726</strong></td>
<td><strong>378</strong></td>
<td><strong>240</strong></td>
<td><strong>3524</strong></td>
</tr>
</tbody>
</table>

Table 5.12: Post-adjustment distribution of NURSE tokens according to following phonological co-texts
Figure 5.10: NURSE plot for realisation patterns based on orthographical taxonomy. The close proximity between Y (older) & O (younger) speakers’ tokens signals the stability of NURSE in apparent time (n=3524)

The weakest of the significant values was however returned for realisation_NURSE ($R^2 = 0.15$ $p=0.0102$). The degrees of freedom for this model were same as in others.

In comparison between the results, a number of striking observations can be made. First, they unanimously confirm the tendency toward NURSE(s) centralisation, hence significantly differentiated from THOUGHT, BATH or DRESS in the system. The effect of realisation_NURSE in the analyses appears to differentiate NURSE from those pairs in the system, therefore suggesting a contrastive realisation. Put differently, the results present a gradient contrast to series of orthographic frameworks thus far identified for NURSE classification in NigE – especially as they relates to fusion with the peripheral classes or utter displacement from mid-central position (cf. Eka 1985, 2004, Odumu 1987, Jowitt 1991, Simo Bobda 2000, Adetugbo 2004). The corroborating graphical evidence in Figure 5.10 however supports a dual phonemic status for NURSE. Similar to Schmied (1991a) and Simo Bobda (2000), the plot confirms the extreme backing of NURSE towards the low back region,
and another realisation somewhere between BATH and DRESS in F1. With regard to individual tokens, while it is plausible to conjecture the cluster of <or, ur and our> as responsible for backing towards the low back region, the extent to which the <ear, ir and er> words approximate to DRESS or BATH remains fogged in the lexical pool or perhaps, impossible to investigate further. For instance, it would be difficult to mechanically attempt the identification of all <ear, ir and er> which might have either been pronounced as DRESS or BATH; or possibly determine what factors might be underlying their randomness. Another universal trend from the analysis is the constancy of mainly linguistic predictors on the list of significant variables. Consequently, the role of age, gender and education were shown as insignificant in the overall achievement of NURSE (cf. Simo Bobda 2000:41). In the light of regional trajectories of the two NURSE phonemes as shown in Figure 5.10, a new set of NURSE vowels re-codable as NURSE-back and NURSE-mid would suffice.

5.2.7 FACE, GOAT and CURE

As explained in Section 2.6, the literature is surprisingly discordant on the diphthongal quality of FACE, GOAT and CURE in NigE. A good number of studies have however either reported the absence or presence of substantive glide (see Jibril 1982, Eka 1985, Odumuh 1987, Jowitt 1991, Awonusi 2004, Adetugbo 2004, Udofot 2004). While some of these reports have equally noted the chance of alternative realisations between /e/ and /o/, /et/ and /oʊ/ or /ou/ for FACE and GOAT, especially by educated/sophisticated speakers, others have reported a pure monophthongal realisation (Adetugbo 2004, Jowitt 1991, Eka & Udofot 1996). The clearest of variationist allusions in the realisation of FACE and GOAT are, perhaps, have been hinted in Schafer (1967) and Afolayan (1968), and recently, Ugorji (2010). In the two earlier studies, the half close vowels /e/ & /o/ are found as constant substitutes for FACE and GOAT, noting further the fusion of /et/ FACE and /eə/ SQUARE in Yoruba accent. Also, the shortening of GOAT to /o/ in go, bone, dope, dose, etc. also frequently results in allophonic similarity with LOT /ɔ/ in Hausa English (Banjo 1993). Ugorji’s analysis derives chiefly on his lectal clines (Section 2.4.2), with a claim of twelve pure vowels for acrolectal and mesolectal varieties, including FACE and GOAT. The fall-out of this is thus the widespread non-distinction between FACE and DRESS – such that the FACE vowel in: day, break, take, bake, etc often coalesce with DRESS in: weight, wet, dead, etc (see also Nuttall 1961 & 1965 on Hausa English).
The inclusion of FACE and GOAT in this analysis is in sync with Hoffmann (2011:165) who cautions against imposing an *a priori* monophthong-diphthong distinction between vowel classes in the description of vocalic inventories. Based on his measurement of glide trajectories in centralised GOOSE (coded as USE in this study), as well as diphthongal strength of FACE and GOAT in BIKE, he canvasses the need for such judgement to only be made after carrying out both visual and statistical assessments. Inspired by Tsukada (2008), the values of F1 and F2 were taken at 20% onset and 80% offset of the vowels and assessed in a linear mixed-effects analysis for the effect of glide movement (Hoffmann 2011:159). Based on these procedures, part of his findings, is the monophthongisation of FACE and GOAT – which also is clearly supported by their lack of significant effect for glide. Compared with PRICE (which is fully diphthongal in the system) neither of FACE nor GOAT shows a visible glide movement on the F1/F2 plot. Therefore, the Euclidean metrics between the vowels’ nucleus/glide in formant one and two were measured while their mean values were evaluated via a two sample *t*.test (Section 4.3). With respect to diphthongs, the Euclidean Distance (ED) represents the Hertz length between the formant values (in F1&F2) from the onset through (20%) the offset (80%) in a two-dimensional coordinate system (cf. Gorman and Johnson 2013:230-4, Hofmann 2015:227). Generally, the diphthongs were measured at 20% for onset and 80% in the offset. Since ED scores are mostly subjective, the post-hoc dependent sample *t*.test was considered suitable for the significance estimation of the metrics.

The ED for GOAT was 0.140. The mean values for $F_1$ & $F_2_{onset}$ were -0.729 and -1.008; and for $F_1$ & $F_2_{glide}$: -0.723 and -0.868. The ED for FACE was 0.092 with mean values for $F_1$ & $F_2_{onset}$ of -0.818 and 1.075; and $F_1$ & $F_2_{glide}$: -0.904 and 1.104. The weakness of glide movement in CURE was reflected with a low ED score of 0.88 with a mean value for $F_1$ & $F_2_{onset}$ at 0.94 and -0.11. The mean $F_1$ & $F_2_{glide}$ were 0.11 and 0.23. Similar to F1 & F2 all analyses in this study, the ED metrics were based on lobanov-normalised formant values. The apparently weak glide for GOAT in Figure 5.10 was corroborated by an insignificant value: $t(-1.17) = 1.061$, $p=0.442$. FACE is evidently monophthongal on the plot, thus insignificant $t(-0.355) = 1.883$, $p=0.758$. The mean difference between CURE’s nucleus and glide also lacked significance $t(0.5071) = 1.0257$, $p=0.699$. These numerical supports regardless, the status of supposedly diphthongs as pure vowels in EEng is further reinforced in Figure 2. The glide extension in FACE, GOAT and CURE demonstrates a much weaker elongation relative to other complex classes.
Some of the missing links in previous descriptions of NigE vowels are: (a) the aridity of standard methodologies for convincing representation of speakers’ vocalic inventories, and (b) the surprising blindness to widespread divergences in speakers’ primary tongues. More specifically, the inventories reported from findings on NigE accents have been mostly based on auditory judgement, devoid of empirical props. Resulting from these subjective accounts is the multiplicity of accounts in the body of literature on the variety. Also, as explained in Section 2.3.1, beneath general commonalities that mark off the NigE system are internal variations – some of which have strong ethnic dimensions (see also Gut, 2004: 815-816).
The goal of analysis in this chapter was, essentially, to foreground these gaps, as well as suggest basic parameters for bridging them. The previous sections have been largely descriptive of the monophthongal inventory in EEng. Being a subset of Nigerian English variety, the scope of analysis was defined by the boundaries of earlier findings on the system, against which the results of this study were weighed.

In the interest of an objective account, the analysis began with no *a priori* assumption on the overall vowel trajectory in the EEng system. However, the categorisation of English accents according to Wells’ Lexical Sets (Wells 1982) was sourced for phonemic coding. Since tokens were collected based on this paradigm, it was hence logical that my analysis (and the review of literature) be presented with the same representation. As outlined in Section 1.2, despite conceptual difficulty associated with classifying homophonous phonemes as ‘mergers’ in NigE, the procedures I employed were mostly consistent with those for assessing mergers or splits on established varieties of English.
The analysis in previous sections was primarily internal – in that only phonological co-texts were considered in token distribution between vowel classes, and the significance of phone label in the statistical assessments. Therefore, the vowel pairs analysed were deemed differentiated or otherwise based on whether the effect of phone label was significant or not. Figure 5.11 depicts a visual summary of the monophthongal inventory for EEng. The assessment of social categories in the manner each of the vowels are realised (see Research Question 2) is discussed in Chapter 6.

Starting with high front vowels, the indication of variance between KIT and FLEECE was supported by a significant effect ($R^2 = 0.10, p = 0.00655$), thus showing these vowels as differentiated in the system. The status of FOOT and GOOSE were assessed in F2 (Gorman and Johnson 2013:232, Hofmann 2015:273) after controlling for the frontish effects of j-words (Mesthrie 2010:10) or the USE tokens. Similar to the high front classes, phone label was least significant ($R^2 = 0.21, p = 0.0144$), led by phone duration and phonological co-texts. Figure 5.5 confirms the full coalescence of LOT and THOUGHT as shown in the regression analyses. The variance between both classes was insignificant for phone label in F2 ($R^2 = 0.273, p = 0.826$). The effect of phone duration was however highly significant, thus indicating a quantitative differentiation. Despite the visual proximity of STRUT with the new class of low back vowel(s), a further comparison showed a differentiation between them. Phone label had the least significant value ($R^2 = 0.10, p = 0.000531$) relative to phone duration ($R^2 = 0.10, p = 3.12e-07$), phonological co-texts ($R^2 = 0.10, p = 0.00546$), and style ($R^2 = 0.10, p = 0.00247$), thereby suggesting a strong quantitative differentiation rather than vowel quality (Figure 5.6).

Due to imprecision on the most suitable outcomes in which TRAP and BATH can be suitably measured, separate models were constructed with F1 and F2. While phone label was statistically significant in the model for F2, its effect was insignificant for F1. Owing to this discrepancy, a follow-up MANOVA and post-hoc ANOVA fitted with: phone label, age group and style as predictors showed no significant effect for phone label in neither F1 nor F2, suggesting a single system for the TRAP and BATH (Section 5.2.5.1). Further evaluation of the low central cluster and LETTER in Section 5.2.5.2 showed a clear differentiation between the classes, especially in F1. The phonemes are visibly separated by phone label and style, especially in height (Figure 5.9 a & b). The measurement of NURSE realisation in Section 5.4 confirms a dual system, one of which is arbitrarily achieved between BATH and DRESS. Compared with THOUGHT in F2 and separately with DRESS and BATH in F1, very high significant values were returned for realisation_NURSE, indicating some extent of NURSE
distinction from the adjacent vowels (cf. Simo Bobda 2000:41, Adetugbo 2004). The analysis however supports a dual system for nurse in the variety – one which is considerably retracted towards the low back classes and the other often achieved between bath and dress; both of which have hence been re-coded as 

_nurse-back_ and 

_nurse-mid_ respectively (Figure 5.10 & 5.12). Though shown in Figure 5.12, the tendencies of use and dress are discussed only in Section 5.3.2.3 and 5.3.2.7 in view of their obviously distinguished status within the entire system. For example, while use is manifestly frontish and separated from goose (see also Mesthrie 2010, Hoffmann 2011), dress also shows clear distinction from neighbouring classes in F1 & F2.

The classification of face, goat and cure as monophthongs was established on Euclidean metrics between their nucleus and glide, and their mean comparison via a two sample _t_-test. For neither of the vowels was the mean difference significant, confirming the weaker glides indicated in Figure 5.10 and 5.12. The phonemes were hence coded as monophthongs for the variety. With the exception of acoustic contrast, an additional detail in the analysis of bike (Hoffmann 2011:164) is his reference to substrate transfer from the speakers’ local pool into their English variety. Following conclusion on the monophthongisation of face and goat, he conjectures the retraction of dress from face and the advancement of cloth from the low back clusters on the inflectional forces of the ATR/RTR dichotomy in the primary systems. The widespread ATR/RTR phenomena in most West African languages (Section 2.3.1) have been widely observed to demonstrate a similar mechanism (Fulop, Kari & Ladefoged 1998: 87–8). While this pattern is conceivable for mid-front and back vowels, it is unclear if the ATR/RTR of the high front and back in Ebira L1 coincides with differentiation already found between kit & fleece, and foot & goose in Section 5.2.1 and 5.2.2.

As reviewed above, many studies on NigE phonology have variously pegged its size of pure vowels at: 5 (Christopherson 1954), 11 (Eka 1985, Odumuh 1987), 7 (Adetugbo 2004), 12 (Ekong 1978, Jibril 1991), 7 (Awonusi 2004) 6 (Eka & Udofo 1996, Udofo 2004), 8 (Jowitt 1991) and 7, 8 and 12 for basilect, mesolect and acrolect (Ugorji 2010). The foregoing supports a 13-vowel monophthongal inventory for EEng. The methodological disparities and speakers’ sociolinguistic differences – all of which often weigh significantly on vowel behaviours however make logical comparisons with any of the previous findings difficult.
5.3.1 Extralinguistic Variables

The previous sections focused mainly on the description of the monophthongal inventory in EEng by assessing the status of all historically non-differentiated vowels in the broader accounts of NigE vowels. To this end, conclusion on the status of each vowel pair was based on the effects of phone label as well as phonological co-texts. As a result, though the non-linguistic variables such as: age-group, gender, number of L1, education degree and style, as well as within-group or inter-speech differences were assessed alongside the linguistic variables, their effects were not explained. The following sections thus substantially discuss the influence of these predictors in the light of some of the vowels in the system.

Socially-concerned investigation of African English varieties has been mainly limited to ethnicity and education (Huber & Brato 2008, Oladipupo & Akinjobi 2015). Considering the dynamics of these emerging varieties, identifying their social definitions could be thorny. For most African countries, socio-economic stratification, e.g., Working Class (WC), Lower Middle Class (LMC), Middle Class (MC) and Upper Class (UC) are usually not as clear-cut as in domains of established English varieties (cf. Labov 1990: 224, Trudgill 1972: 91). However, due to ethno-linguistic origins and disparate access to formal education, the role of ethnicity and education have been recurrently prominent in the taxonomy of NigE phonology (Brosnahan 1958, Banjo, 1971, Bamgbose 1982, Udofo 2003, Gut 2004, Olaniyi 2014).

Given the concept of speech communities – which involves speech cultures formed on network relations as well as idiolectal divergences (Lesley Milroy 1980, 1981, 1987, Mufwene 2003), such broad classifications would be hard to explain. For instance, the spectrum of ethnicity includes a laundry of confounding factors, some of which are independent of speakers’ nativity. Similarly, the North/South delineation often implicated in the construction of ethnicity has been called to question, especially in the context of Nigeria’s geo-political and linguistic multiplexities (cf. Jibril 1986, Ugorji 2010). Another is the concept of education which remains blurred – as far as speech proficiency is concerned in L2. The assessment of correlation between speakers’ phonological behaviours and their level of education also becomes problematic in view of other concomitant variables, including childhood neighbourhood and the speakers’ nascent or domestic environment. Thus, a speaker’s style of speech can be a hybrid of multiple voices from an array of linguistic exposures – such that the effect of education, for instance, is effectively stringed with other
social variables. The assessment of these ‘hidden’ predictors have, therefore, been included and discussed in this section.

With regard to age cohorts, the Labovian construct of age-grading and other well-known parameters were practically constrained. The preliminary test for correlation between speakers’ real time age, exposure to L2, years of exposure to L2 and vocalic behaviours indicated very weak effects, and as a result, excluded from regression models. From the outset, a post-retirement group (comprising of speakers above 56 years) were excluded due to difficulty in recruiting speakers who meet the minimum academic qualification for population. Theoretically, speakers of this age range often tend to loosen up on the conservative style, being already disengaged from the ‘marketplace’ (Labov 1972, Keith 1980). Among NigE speakers however, especially for EEng, very few speakers in this cohort would have had tertiary school certificates or might speak English fluently. As a replacement for the older generation, respondents who were through the middle of their career age up to the eve of retirement (35 – 56 years) were recorded. This age-bracket is mostly linked to linguistic conservatism much likely inspired by workplace pressures and institutional social networks (Sankoff & Laberg 1978, Milroy 1980, Edwards 1992, Eckert 1996). The decision to include two speakers: FUO4 and MUO7, both 31 and 34 was anchored on observed in-group semblance between them and speakers above 35.

The younger age group, those in the range of adolescence (15-18) were excluded. Speakers at this stage are mostly prone to linguistic changeability – arising from exposures to diverse peer norms (Eckert 1998, Hofmann 2015). Their exclusion was also reinforced by the official entry year into Nigerian tertiary institution which is currently pegged between 16 - 18 (though it is possible to find undergraduates of a lesser age), and since my aim was to somewhat control for education by including only speakers in tertiary schools or those who had tertiary school education, it was harder to recruit respondents from this rank. My younger generation of speakers however included who, instead of the public elementary schools, attended crèche or private nursery schools, most of whom were taught only in English right from the outset. From both gender, therefore, the speakers in this group were mainly between the ages of 19 – 25. Two speakers, MUY3 & MUY6, 28 and 26 years were however included based on demonstrable in-group equivalents.

Excepting conceptual complexities which often trail gender taxonomy, speakers’ gender was determined on the traditional grouping of male and female (Labov 1990:221-2). As discussed in 3.2.2, access to formal education by female children in Ebiraland was widely
unpopular until about three decades ago. Seen as potential housewives whose core duties
would be mere house-keeping and baby-making, their social engagements was culturally
inhibited. This trend, however, is being gradually reversed, particularly since the last two
decades – hence the prospect of further effect for age among female speakers. At the design
stage of my study, the speakers’ level of education was intended to be uniform, such that all
participants would have either bagged a tertiary school certificate or would be in the process
of getting one, so that only ‘educated speakers’ were recorded (cf. Brosnahan 1958, Banjo,
1971, Bamgbose 1982, Udofo 2003, Gut 2004). Things were however not as straightforward
on the field. The total of 11 speakers had university degrees, 14 had Higher National
Diploma, 2 with the National Certificate in Education (NCE), 2 had Masters in Science
(MSc), and 1 with a Postgraduate Diploma (PGD). NCE was merged with HND and PGD
with MSc, while BSc was retained as a separate category. For statistical analysis, the subsets
were broadly grouped into HND and BSc holders, considering the fact that speakers with
PGD and MSc certificates had also bagged a BSc degree. Though mostly subsumed as
‘educated speakers’ in prior studies, I decided on splitting them for possible effects (if any) –
of the psycho-social dichotomies furtively etched on them in Nigeria.

Other included variables were speech styles and the number of first language(s). The
evaluation of speakers’ behaviour through different speech styles has become particularly
important since Labov’s assessment of r-realisation across different speech styles in New
York City (Labov 1966:240). Following the notion of style-induced variation in speech, the
gradience of realisation across the styles was also statistically examined. Though only
speakers who had lived for most part in Ebiraland since childhood were recorded, I was
curious to learn if some speakers equally have learned or acquired another Nigerian language.
The effect of number of L1 was however insignificant in all previous analysis, hence excluded
from the statistical analyses in this section. Since the study only involved Ebira speakers,
ethnicity was not considered.

5.3.2 Regression Analysis

The models for gauging the social effects basically shared similar principles with the structure
already used in weighing the extent of fusion or differentiation between the tested clusters. In
order to ensure fair distribution between the vowel classes, their tokens were cross-tabulated
with phonological co-texts. Levels with comparatively fewer tokens were either combined
with phonologically similar groups or excluded from the data before regression runs. The data structure for each cluster – as ‘adjusted’ for previous analysis was retained. In view of the research question however, the analysis of each vowel class was separately done – such that ‘phone label’ was automatically excluded from the list of predictors in the models. Essentially, the data structures were split according to each vowel before cross-tabulation and regression runs.

5.3.2.1 KIT and FLEECE

The regression results for the high front vowels in Section 5.2.1 generally support the appropriacy of assessment in F1 with the complex model. Results from this model showed a general differentiation between the pair by: phonological co-texts, gender, and phone label. To further examine if similar outcomes would be obtained for KIT across the social groups, only the KIT tokens were retained for analysis. This automatically excluded ‘phone label’ from the model – since a single observation (KIT) left under it could not be evaluated against itself. Aside phone label and Number of L1, other fixed predictors were as in previous models, with individual speakers and word items for random effect. To measure the effects of age and gender, an interaction was created between the two variables and entered as a single predictor. The total number of analysed tokens was 1070. As in the analysis in Section 5.2.1, results for KIT showed the strongest significance for phonological co-text ($R^2 = 0.11$ $p= 1.9e-05$), followed by phone duration ($R^2 = 0.11$ $p= 0.0224$) and age*gender ($R^2 = 0.11$ $p= 0.0412$). The factors of: education degree and style were insignificant. The obvious divergence between this run and the previous analysis of high front was only in phone duration (which means that all other predictors behaved the same way as in the previous model). The disparity between realisation patterns was again marked for gender across the age groups, thus reinforcing the differentiating effects of age and gender between the high vowels.
Table 5.13: Significant predictors in Kit realisation \((R^2.\text{fixed}=0.087 \quad R^2.\text{random}=0.268)\)

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
<th>(R^2)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonological co-texts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>velars</td>
<td>0.141</td>
<td>272</td>
<td>-0.654</td>
<td>(R^2=0.36) (p=1.9\times10^{-5})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>labio-dentals</td>
<td>0.020</td>
<td>646</td>
<td>-1.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatals</td>
<td>-0.121</td>
<td>152</td>
<td>-1.153</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phone duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuous +1</td>
<td>-0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age*gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEM.OLDR</td>
<td>0.072</td>
<td>278</td>
<td>-0.858</td>
<td>(R^2=0.36) (p=0.0412)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEM.YNGR</td>
<td>0.046</td>
<td>210</td>
<td>-0.966</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE.YNGR</td>
<td>0.012</td>
<td>303</td>
<td>-0.942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE.OLDR</td>
<td>-0.130</td>
<td>279</td>
<td>-1.151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker (random)</td>
<td>Std. Dev. 0.129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(Deviance = -493\) \(df = 12\) \(Mean = -0.979\) \(AIC = 1058\) \(Overall intercept = -0.027\)

Figure 5.13: Graphical illustration of gradient differences in Kit realisation between gender speakers and age–groups (\(n=1070\); speakers = 29)
The procedure for FLEECE was same as for KIT above. Phone label and the number of L1 were not included in the model, while age and gender were again put together as one variable for a combined effect. Other fixed effects comprised: education degree, style, phone duration and phonological co-texts, with individual speakers and word retained as random effects in F1. Age*gender yielded the strongest significance value ($R^2=0.11$ $p=0.000129$). As in the previous analysis, phonological co-text was also significant ($R^2=0.11$ $p=0.0263$), and the marginal effect of education degree ($R^2=0.11$ $p=0.0478$). Phone duration was however not significant in the case of FLEECE ($R^2=0.11$ $p=0.865$). The AIC was 542 while the deviance was -226, with the overall intercept and means of -1.154 and -1.141 on 12 degrees of freedom.

The major parallels in the analysis of the high front pairs, especially across age and gender groups are socially instructive. While the lone effect of age was insignificant for both vowels: KIT ($R^2=0.10$ $p=0.238$) and FLEECE ($R^2=0.11$ $p=0.395$), high significant values were returned for the combined effects of age and gender, indicating a cline of differences between the groups. Owing to the paucity of theoretical positions with regard to effects of gender and age on high vowels, it is difficult to explain what such interaction indicates or forecasts for the variety or NigE in general. For instance, it would be tricky to assert the movement of each vowel for either age or gender against a previously established pattern of behaviour in NigE speakers – given the absence of such record in the literature. Another remarkable pattern was the effect of education degree for FLEECE. The result revealed a marginal, yet significant difference between speakers with Higher National Diploma and Bachelor of Science degrees ($R^2=0.11$ $p=0.0478$). The prospect of this variation nevertheless appears weak and may not be considered strongly suggestive, especially in view of its occurrence only in one of the high front clusters.

Regrettably, studies on the structure of the high front vowels in established varieties of English are still relatively sparse. And despite tons of descriptions of these vowels for many African varieties, their trajectories – as defined by speakers’ age and gender are still awaited. For example, neither of the two acoustic evaluations of African systems reviewed in this study includes explanation on gender and age effects in their findings (Mutonya 2008:446, Hoffmann 2011:162-3). Mutonya particularly forecloses the chance of measuring these factors by stratifying his data chiefly on broad national boundaries, with no assessment of the gender groups in his data.
Though Hoffmann’s data for Black Kenyan English appears to have been sourced with the speakers’ age in mind, its effect is unattended in subsequent statistical assessments. Despite the comparative shortfalls, the effects confirmed for this variety steadily supports studies on other varieties in which age and gender have been identified as strong markers of between-group differences (Labov 1990, Brato 2012, Hofmann 2015).

5.3.2.2 GOOSE and FOOT

Unlike for the high front classes, some two major outings – one of which has a strong sociophonetic orientation, recently examined the shades of GOOSE vowel for South African and Kenyan varieties (Mesthrie 2010, Hoffmann 2011). Depending on the phonological environment, the GOOSE phonemes are inherently ‘frontish’ or ‘backish’ (Section 5.2.2). As discussed in 5.2.2, most GOOSE tokens in preceding /j/ context (Figure 5.3) noticeably tend towards fronting. In fact, some of such tokens often occupy the same region as the high front vowels, such that they are more or less front rather than frontish. Other consonants with either advancing or retracting effects on GOOSE are coronals and alveolars (Lanham 1978:152, Mesthrie 2010:10). These two studies however differ in reference and methods. While Hoffmann’s findings involve a cursory behaviour of GOOSE in Black Kenyan English, Mesthrie investigates a range of structural inroads into the vowel with regard to speakers’ race and gender extraction in major South African accents. For instance, drawing on a proposed ratio by Watt & Fabricius (2003) for assessing the movement of high vowels, he finds a strong leaning towards fronting among the White population for both coronal and non-coronal environments. The overwhelming force of this pattern, he notes, is also budding among some other groups of speakers. Though the pattern is primarily typical of mostly Whites South Africans, the trend also suggests a gradual accommodation of the Black Middle Class South Africans and the youths, with the female speakers much farther on the wheels of change (Mesthrie 2010:28). The sociolinguistic parallels of these definitions in other African varieties are therefore open to further investigation.

With regard to consonantal influences in current data, GOOSE tokens are relatively frontish in preceding palatal and backed in velar environments (Figure 5.4). As in other systems, clear front status is similarly marked in the variety, with proneness to high front configuration (Figure 5.3). The variation in GOOSE realisation was largely defined by education degree ($R^2=0.15$ p=0.0164). Other significant predictors were internal, including
phonological co-texts ($R^2=0.15$ $p=0.0217$) and phone duration ($R^2=0.15$ $p=0.0257$). The differentiating effects of gender and age were nonetheless insignificant ($R^2=0.15$ $p=0.273$) and ($R^2=0.15$ $p=0.446$). The $R^2$ random and fixed values were 0.59 and 0.15 respectively.

Similarly, the analysis of FOOT yielded statistical significance only for age-group ($R^2=0.21$ $p=0.00505$) among other social variables. The combined effect of age and gender was equally significant ($R^2=0.21$ $p=0.0448$), with $R^2$ fixed and random values of 0.21 and 30 on 11 degrees of freedom. Education degree however showed no significance for neither of the vowels as their between-group variance appeared to have been more driven by age rather than education degree – at least in F2 (Appendix C.2).

### 5.3.2.3 USE

The aridity of descriptive accounts on the status of USE in NigE, might perhaps, been strongly anchored on subsisting claims of yod-deletion in all C_/u/ environments (Simo Bobda 2007:288). Later studies on NigE vowels seem to have so far dismissed the likelihood of USE occurrence in the system, since going by Simo Bobda’s report, ‘yod-deletion applies rather inordinately’, and ‘nearly systematically’. As far the literature in NigE is concerned, and as much as I am aware, the boldest attempt on the assessment of the analysis of yod is in Oladipupo (2015). Unfortunately, the scope of his study was delimited to surveying the prosodic planes of yod coalescence across word boundaries. In the absence of any known critic of Simo Bobda’s theory, the USE vowels are generally assumed as absent in typical NigE inventories.Interestingly, the evidence so far in the current data signals a clear differentiation of post-yod GOOSE tokens (coded as USE) from others in non-yod environments (Figure 5.1, 5. 3 & 5.11). While the question of phonemic differentiation between USE and GOOSE in these varieties needs not much discussion, it is significant to measure its internal trajectories among the social groups.

For USE analysis, tokens were distributed according to following and preceding environments, comprising occurrences in palatal, apical and labio-dental co-texts. F2 was chosen as the dependent variable, with phonological contexts, phone duration, age group, and gender, education degree as fixed effects; and word label as well as individual speakers as random effects. Unsurprisingly, the strongest effect was shown for phonological co-texts ($R^2=0.20$ $p=0.0024$), then phone duration ($R^2=0.20$ $p=0.0124$). Again, only age-group was socially differentiating between the groups ($R^2=0.20$ $p=0.0477$) with a mean difference of
0.324 between the older (0.189) and younger (0.513) generations, indicating a more frontal realisation for the younger speakers. The discrepancy between the approximate $R^2$ fixed and random values was rather minimal at 0.20 (20%) to 0.30 (30%). Given this strong effects of the explanatory variables, the evidence of age-conditioned variation is thus reinforced, while in F1, the role of gender, especially for older speakers, is also marked. Since the analysis was run in F2, the insignificant value returned for gender in this dimension was least surprising ($R^2= 0.20 \ p=0.968$), but given the fittingness of USE measurement in F2 (Gorman & Johnson 2013:232, Hofmann 2015:244), the analysis was mainly carried out in the second formant. A rather strong significance of gender is suggested in F1 ($R^2= 0.15 \ p=0.000771$). The effects of other predictors including education degree and style were insignificant.

5.3.2.4 THOUGHT and STRUT

As highlighted in 5.2.1, the LOT and THOUGHT vowels are qualitatively non-differentiated, but as separate classes by length, i.e., phone duration. The fusion of these vowels in EEng is thus indisputably consistent with findings on many other varieties of English in which the vowels have been found to have singular phonemic status (Wetmore 1959:115, Irons 2007, Hofmann 2015). With respect to social gauging of this behaviour for Newfoundland variety, Hofmann (2015:264) finds the merger situation to be generally stable in apparent time – which implies an overall lack of effect between the age groups. Though none of the previous reports on these vowels for NigE has considered the effect of age, the regression results in 5.2.1 returned a non-significant value for age-group, thus corroborating Hofmann’s findings. In clear concurrence with tense/lax pattern for most NigE varieties, the analysis showed a high significance for phone duration ($R^2= 0.05 \ p=7.4e-09$), which going by Klatt (1976), may as well qualify as valid indication of vowel differentiation (Table 5.10).

Following the lack of qualitative distinction between LOT and THOUGHT, both vowels were re-assigned to a single class of THOUGHT and compared with STRUT in Section 5.2.4. The treatment of STRUT as a member of the low back cluster was premised on several subsisting claims of extreme retraction of the vowel, such that it becomes identical with either of lot or THOUGHT (Jowitt 1991, Eka & Udofo 1996, Collins & Mees 1999: 23, Udofo 2004, Awonusi 2004, Fabricius 2007: 296). While the subsequent evaluation of STRUT against THOUGHT (Table 4.14) showed a very high durational difference ($R^2=0.10 \ p=7.03e-09$),
### THOUGHT

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>phone duration</td>
<td>-0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>formal</td>
<td>0.028</td>
<td>418</td>
<td>-0.026</td>
<td></td>
</tr>
<tr>
<td>informal</td>
<td>-0.028</td>
<td>1233</td>
<td>-0.988</td>
<td></td>
</tr>
<tr>
<td>Speaker (random)</td>
<td>Std. Dev. 0.111</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Deviance} = -117 \quad df = 13 \quad \text{Mean} = -0.998 \quad \text{AIC} = 321 \quad \text{Overall intercept} = -0.982 \]

### STRUT

<table>
<thead>
<tr>
<th>predictor</th>
<th>coefficient</th>
<th>tokens</th>
<th>mean</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonological co-texts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>palatals</td>
<td>0.112</td>
<td>519</td>
<td>-0.735</td>
<td></td>
</tr>
<tr>
<td>velars</td>
<td>0.020</td>
<td>81</td>
<td>-0.869</td>
<td></td>
</tr>
<tr>
<td>apicals</td>
<td>-0.038</td>
<td>245</td>
<td>-0.971</td>
<td></td>
</tr>
<tr>
<td>labio-dentals</td>
<td>-0.094</td>
<td>898</td>
<td>-0.986</td>
<td></td>
</tr>
<tr>
<td>phone duration</td>
<td>-0.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>formal</td>
<td>0.035</td>
<td>288</td>
<td>-0.851</td>
<td></td>
</tr>
<tr>
<td>informal</td>
<td>-0.035</td>
<td>1455</td>
<td>-0.915</td>
<td></td>
</tr>
<tr>
<td>Speaker (random)</td>
<td>Std. Dev. 0.136</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Deviance} = -767 \quad df = 13 \quad \text{Mean} = -0.904 \quad \text{AIC} = 1607 \quad \text{Overall intercept} = -0.879 \]

Table 5.14: Results of significant variables in the analysis of THOUGHT and STRUT, excluding the number of L1 and phone label: THOUGHT \(R^2=\text{fixed}=0.051\quad R^2=\text{random}=0.272\), STRUT \(R^2=\text{fixed}=0.077\quad R^2=\text{random}=0.209\)
phone label was least significant ($R^2 = 0.10 \ p=0.000531$). Weighed against previous reports of homophonous overlaps between these classes, as well as my auditory assessment of the speakers during our interview sessions, a further investigation into the odds of by-speaker idiosyncrasies would be in order – especially given the graphical proximity between STRUT and THOUGHT as revealed in individual speaker’s pattern (Appendix D.3.2). For example, barely 3 of the 29 speakers appear to demonstrate a unique pattern of realisation for these classes, and these speakers belong to the broad category of younger groups. A leaning towards homophonous overlap was fairly maintained by the older age bracket.

The procedure for the investigation of THOUGHT and STRUT was the same as in 5.2.4. Both vowels were assessed separately in F2 with gender, age-group, education degree, style, phonological context and phone duration as entered as fixed factors. Individual speakers and word items were retained as random predictors. Table 5.14 shows the results of significant predictors for the vowels. The drivers of variation in both vowels were consistently phone duration and style, while the effect of phonological co-texts was only significant for STRUT. The exclusion of phone label and the number of L1 apparently had no implication on the outcome of results, since the outcome was quite similar to the previous runs. Importantly, the repeated significance for style in the runs unarguably lends further credence to Labov (1990: 224), especially on the propensity towards style-shifting. Based on these analyses, variation in the low back groups are driven by mainly linguistic variables, namely: phone duration, phonological contexts and style. In contrast to the high vowel classes, the lack of age-conditioned differences for THOUGHT and STRUT indicates their stability in apparent time (Hofmann 2015:249).

### 5.3.2.5 BATH and lettER

Section 5.2.5.1 discusses in detail the procedures and results leading to the re-assignment of TRAP and BATH classes into a single class of BATH for this data. The results thus provide strong support for some earlier reports of non-differentiation between the two classes (Eka & Udofot 1996, Adetugbo 2004), as well as durational differentiation (Jibril 1982, Awonusi 2004). Interestingly, these behaviours seem to agree with patterns that have also been attested for some varieties of Scottish and Canadian English (Nycz & Hall-Lew 2014:11-4, Hofmann 2015:253-5). Given this consistency, the phonemic identity of both classes was re-coded as BATH before comparing it with lettER in Section 5.2.5.2. The difference between BATH and lettER was however very significant ($R^2 = 0.31 \ p=2.14e-12$) in F1 (cf. Jowitt 2015:42), thus
distinguished from BATH (Figure 5.9a). The role of speech styles in the separation of these vowels showed an equally high significance ($R^2 = 0.31 \ p = 6.54 \times 10^{-10}$).

The tokens of BATH and letter as distributed according to the following co-texts were assessed in separate regressions. Age-group, gender, education degree, style, following co-texts and phone duration were entered as fixed factors with individual speakers and word items as random variables. The analysis showed the effect of style ($R^2 = 0.17 \ p = 1.69 \times 10^{-11}$) as most significantly differentiating for BATH. Other significant factors were phonological co-texts ($R^2 = 0.17 \ p = 0.0186$) and phone duration ($R^2 = 0.22 \ p = 0.0301$). None of the key social predictors were shown to have further significant effect. With the same list of predictors, only the effects of phone duration ($R^2 = 0.15 \ p = 4.68 \times 10^{-5}$) and style ($R^2 = 0.15 \ p = 0.000886$) were most significant in letter. The AIC for BATH was $-437$ with the $R^2$ fixed and random at 0.17 and 0.19 on 12 degrees of freedom, and letter at 342, with $R^2$ fixed and random = 0.158 and 0.303 on 12 degrees of freedom. Since statistical assessment of these variables was only in F1, it was unclear what the effect of a gender-induced difference signalled in Figure 5.13 (mainly in formal speech styles) for both vowels along F2 might suggest. Similar to outcomes for low back classes, variation in BATH and letter was driven predominantly by linguistic factors, especially by the overarching effect of style and phone duration. The importance of phonological contexts in BATH was however not repeated for letter, possibly as a consequence of its usual phonological environment, i.e., in word final positions (Wells 1982). This co-textual immobility would logically impose a weakening effects on the statistical outcome of co-texts ($R^2 = 0.15 \ p = 0.352$).
Figure 5.14: Plot showing the trajectories of differentiations in BATH and letter – by gender and style (n =1030)

5.3.2.6 NURSE

The analysis in Section 5.4 generally shows the tendency towards a dual system for NURSE. Ostensibly, the results confirm a clear contrast between NURSE in <ear, ir & er> and NURSE_DRESS or NURSE_BATH as well as between NURSE in <or, ur & our> and NURSE_THOUGHT, thus suggesting a two phonemic possibilities. Consequently, some indication of general progression towards NURSE centralisation in the system and a palpably two phonemic system was attestable in the data, i.e., NURSE-back and NURSE-mid.

As mentioned in Section 5.2.6, while it was logical to conjecture the backing of NURSE-back on the achievement of <or, ur & our> orthography in lexical items, the theory is not as clear-cut for NURSE-mid. For instance, it was difficult to determine the scale of <ear, ir & er> words realised as BATH or DRESS in the data. Beyond the odds of word-specific variation, such haziness might also have some social implications. In fact, apart from
orthography, the major differentiating parameters for NURSE are marked as largely involving education, age group, ethnicity or geolinguistic zones (Simo Bobda 2000:41).

Simo Bobda’s notion of speakers’ age and their level of education is desirable, since the variables can be concurrently tested in the current data. It is however important to note the discrepancy that exist between his construction of ‘education level’ and my ‘education degree’ in this study. While the former entails categories of school leavers at different stages, e.g., elementary, secondary and tertiary levels, those involved in my study involved only tertiary school degree holders or those in the process of getting such degrees. In what follows, I analysed the two identified subsets of NURSE so as to verify their extent of variation among the groups.

For NURSE-\textit{back}, age-group, gender, style, phonological co-texts, phone duration, education degree were fitted as fixed variables with speakers and word items as random effects. Since progression for this vowel is towards backing, F2 was held as the dependent variable. Style had the highest effect ($R^2 = 0.12$ $p = 0.000633$), followed by phone duration ($R^2 = 0.12$ $p = 0.0177$) and gender ($R^2 = 0.12$ $p = 0.0338$) on 11 degrees of freedom. The overall mean values and intercept were -0.72 and -0.76 respectively, with AIC at 385 and a deviance of -166. The $R^2$ values for the random and fixed predictors were 0.11 and 0.22. Though the same independent variables were considered in the analysis of NURSE-\textit{mid}, the outcome variable was F1, since the vowel appears most significantly defined in the height dimension. The effect of style again yielded the strongest significant value ($R^2 = 0.15$ $p = 0.000289$). The two other influential factors similarly include phone duration ($R^2 = 0.15$ $p = 0.00896$) and gender ($R^2 = 0.15$ $p = 0.0492$) on 11 degrees of freedom. The AIC was 385 with grand intercept and mean at 0.58 and 0.87. The $R^2$ values for fixed and random variables were 0.15 and 0.47.

Compared with the results in Section 5.4, a commonality could be established. The effects of mostly internal factors have remained consistent for this vowel. For previous analyses involving two different classes, results generally showed a strong tendency towards differentiation by speech styles, phonological contexts and phone duration, thus suggesting lack of effects for social factors in \textit{realisation}_NURSE. In the current analysis, however, marginal significant effects were additionally shown for gender in NURSE-\textit{back} ($R^2 = 0.12$ $p = 0.0338$) and NURSE-\textit{mid} ($R^2 = 0.15$ $p = 0.0492$). With the exception of this variable, it would be logical to again state the differentiating influence of mainly linguistic factors within the vowels. Correspondingly, the outcomes for this data diverge from the drivers of variation in NURSE as suggested in Simo Bobda (2000:41-2). The effect of age-group ($R^2 = 0.12$ $p = 0.316$)
and education degree ($R^2=0.12$ p= 0.902) were insignificant for nurse-back and also for nurse-mid: age-group ($R^2=0.15$ p=0.29) and education degree ($R^2=0.15$ p=0.854). As explained in 5.2.6, the likely divergence between my constitution of speakers’ level of education and Simo Bobda’s makes it difficult to establish a parallel between the two findings. However, the lack of significance effect for age-group in both analyses suggests a stability of the vowels in apparent time, at least in this system.

5.3.2.7 DRESS

The DRESS phoneme is conventionally achieved as short, lax, front unrounded vowel, often transcribed as [ɛ] or [e] (Wells 1982a:128, Cruttenden 2001:110). DRESS-realisation has also the option of an added schwa-glide, i.e., $[\acute{e}]$ or $[\acute{e}]$ in some ‘affected’ accents of RP (Hughes, et al. 2005:48). While this may structurally be the case, possibilities of age-driven variation as well as geographical factors have been reported (Hawkins & Midgley 2005, Torgersen & Kerswill 2004, Kamata 2008). In a related study of DRESS on Hawaiian English, additional factors involving gender and speakers’ penchant towards Hawaiian Creole are also included as socially defining (Drager et al. 2013:41). For Southern NigE varieties, DRESS is adjudged homogenous, but attested as [ɛ] or as central [a] for Hausa English (Eka and Udofo 1996). The claim of homogeneity in Southern accents has however been disputed in Schafer (1967:13-14) who notes the tendency towards substitution of DRESS with FACE and vice versa by some speakers of Eastern origin. The cases of DRESS-lengthening reported in Ekong (1987) and Jibril (1995) lack corresponding social definitions.

The inclusion of DRESS in this study was initially informed by intent to document all monophthongal possibilities in the system (Section 5.3.2.7), as well as subsequently determine its social routes among the speakers. Apart from its comparison with some tokens of nurse_dress in 6.2.5, the analysis has so far excluded DRESS. This is because of its evidently distinct status as revealed on series of graphical appraisals (Figure 5.1 & 5.2), hence the pointlessness of discussing its presence or otherwise in the system.

Since the vowel has not so far been independently explained, all tokens of DRESS were extracted and examined in the following co-texts. Phonological co-texts, age-group, gender, education degree, style and phone duration were fitted as fixed effects, with speaker and word item as random variables. Due to imprecision in the literature as to the most robust dimension in which DRESS might be measured, I ran separate regressions with F1 and F2 (Kamata 2008
and Drager 2013). In addition, the prospect of comparing both fits for suitability would be attractive. The significance table in F1 was led by phone duration ($R^2= 0.10 p= 6.07e-07$). The effect of style ($R^2=0.10 p=6.57e-05$) and phonological co-text ($R^2=0.10 p=0.00141$) were also significant. Neither of age-group nor gender yielded significant values ($R^2 = 0.10 p= 0.198$) and ($R^2 = 0.10 p=0.299$). The model’s deviance and AIC were -759 and 1595 with the intercept at 0.094. The $R^2$ value for the fixed and random effects returned 0.10 and 0.24 on 12 degrees of freedom. The grand mean for the run totalled 0.077.

The effect of phone duration was again returned as most significant in F2 ($R^2= 0.04 p=0.000206$). A marginally significant variable, education degree ($R^2=0.04 p=0.0518$) was however added to the effect of style ($R^2=0.04 p=0.000541$) and phonological co-text ($R^2=0.04 p=0.0215$). Similar to the previous run, the effects of age-group and gender were again non-significant. The AIC (1542) and deviance (-733) for this model were much closer to the values for F1. The $R^2$ values for the fixed and random effects were 0.038 and 0.312 with the intercept at 0.61 and a total mean of 0.62. While both models appear comparatively well with the data, the AIC value for F2 was slightly smaller than for F1, while its $R^2$ total apparently suggests a better fittingness to the amount of variation explained in the data. As earlier hinted, however, the interpretation of significance for education degree in F2 requires some caution, especially in the light of the mean difference between the BSc (0.65) and the HND (0.59) groups which approximates only to 0.1. With the exception of this variable, therefore, the results from both models were much similar; and either of F1 or F2 would be suitable for assessing this vowel.

### 5.4 Social Differentiation in EEng Vowels

The preceding sections have been mainly focused on the analysis of socially-conditioned variation in the realisation of vowels in EEng monophthongal system. In line with research question 2, the outcomes thus explain how significant the relationship between the social groups and vowels’ realisation – and particularly the major predictors responsible for such behaviour. The separate mixed effect modelling of KIT and FLEECE suggested a high significance for co-texts and phone duration, including age and gender groups, with the additional effect of education degree for FLEECE. Though the effects of all relevant predictors were measured against each of both vowels in separate regressions, the evaluations were
unanimous in the effect of age, gender and phonological co-texts. The divergences however, concerned the role of duration and education degree; respectively for KIT and FLEECE. While these outcomes find no parallel or any sound basis of comparison in previous studies on NigE, the overall influence of age, co-texts and gender in both high vowels corroborates Hofmann (2015:340) on the role of age, gender and co-texts in the achievement of KIT among Newfoundlanders. Based on the individual average for KIT (-2.10) and FLEECE (-1.14) in previous analysis, the vowel appeared raised for younger speakers at a mean value of -1.13, and lowered for older speakers at -1.51. A similar movement was repeated for FLEECE as per the effect of age and gender. With the high vowels, the effect of education in FLEECE was equally reflected for GOOSE, but not for KIT or FOOT. Such asymmetry perhaps, indicates an active interaction between phone duration and education degree, but this was not statistically assessed due to variable dissimilarity; i.e., continuous and binary levels.

Results for GOOSE and FOOT were much akin to the high front counterparts. Also, differences in GOOSE realisation were mostly driven by age, co-texts and phone-duration. The significant effects for GOOSE comprised education degree, phonological co-texts and phone duration, while the analysis of FOOT yielded significant values only for age and gender. The pattern in USE however resembles that of high front vowels with variation mainly informed by co-texts, duration and age groups. In spite of the difficulty in relating these outcomes with cognate findings, the F2 differences in FOOT and USE between older and younger speakers demonstrate a broad correspondence with Mesthrie’s report of high back fronting among White South Africans, Black Middle Class and younger speakers (2010:28).

The analysis of the low back cluster in 5.2.1 resulted in re-assignment of low back vowels into a single class of THOUGHT. The vowel was hence paired with STRUT – which appears distinctly backed for the variety. In either of the measurements, the results were much same as in Section 5.2.2, showing very high significance values for co-texts, phone duration and style; as well as indicating a sustained influence of chiefly internal factors for these vowels. This pattern likewise held for the low central group of BATH and letter. The subsets of NURSE, i.e., NURSE-back and NURSE-mid varied only by style, duration, as well as gender. With the exception of ethnic and geolinguistic categories – which could not be accessed with this data, the remaining effect of speakers’ age and education were insignificant. DRESS also appeared to be very stable in the system with variation only by duration, style and co-texts.

In the light of the foregoing, while age, gender and education degree were socially determining for high and mid vowels, other classes were defined by mainly linguistic
predictors, e.g., co-texts, phone duration and style. As would be expected, linguistic predictors were consistently significant for all the vowel classes (See also Nycz & Hall-Lew 2013:4, Hofmann 2015:228). As explained in Section 5.1, the phonological environments were broadly classified according to a conventional framework; hence the difficulty of defining vowels against specific hunches or attested behaviour (cf. Mesthrie 2010, Thomas 2011, Hofmann, 2015).
6 Conclusion

This study has primarily investigated the monophthongal inventory in Ebira English system, especially in the context of the broader NigE systems, e.g., Yoruba and Hausa English or Southern and Northern English. Second, the prime effects of variables including age, gender, education, style, linguistic contexts, and phone duration were also examined. As pointed out in Section 2.7, though the literature was very resourceful in instructing my research questions, the methodology employed was draws on basic framework for variationist designs – with specific references to the Labovian paradigm. For example, although the instances of vowel coalescence or differentiation in this study are incongruent with conceptual background for vowel mergers and split in variationist literature, the methods I employed for assessing vowels’ status in the system were consistent with those for mergers (Eberhardt 2008, Brato 2012, Nycz & Hall-Lew 2014, Hofmann 2015). Thus, the data were extracted from naturally collected speech tokens – as in Labov, Ash & Boberg (2006:33-5) & Tagliamonte (2006), particularly as it concerns eliciting unaffected conversation or casual speech. By and large, the methods were akin to the traditional routines for analysing mergers and splits.

As part of my objectives (RQ1b), I was also motivated to identify which of Yoruba and Hausa English inventory would demonstrate the closer similarity with the EEng system. Historically, the relationship between the Ebira people and these two major ethnics bestrides linguistic and cultural planes. While the ranks of earliest teachers in Ebiraland were predominantly of Southern extraction, i.e., Yoruba speaking instructors, the mainstream of Ebira nation has, in subsequent years, become more inclined to Northern mores and geopolitical aspirations. This is possibly why 80% of Ebira folks, including my respondents, also speak either of Yoruba and Hausa or both. Investigating the vowel inventory in EEng thus became attractive, particularly in view of underlying phonological differences in these L1s (Section 1.1). My second research question involves further measurement of variables included in the statistical assessments. As discussed in Section 2.4, predictors such as age, gender and the linguistic factors were also assessed.
6.1 Summary and Discussion

This study supports a 15 monophthongal system, including FACE, GOAT and CURE which are evidently without glide in the variety (see also Simo Bobda 2003: 23, 2007:412 & Jowitt 2015:42). Were the trio of FACE, GOAT and CURE to be excluded, the results would generally correspond to accounts of a 12 monophthongal system earlier reported in Ekong (1978), Jibril (1982) and Ugorji (2010). With regard to high vowel classes, the analysis of KIT & FLEECE and GOOSE & FOOT revealed marginal differentiation with the effect of phonological co-texts as the strongest in the models. Though the vowels belong to the historically coalesced pairs in NigE systems (Section 5.0), the indication of contrast between them is consistent with the perceptual account of distinction in Educated Hausa English (Jibril 1986, 1995, Gut 2004). Since age-group was insignificant in differentiating between these vowels, a correspondence with change from the use of indigenous language to English at the lower primary since the mid 1990s was not signalled (Section 2.1.1). If this was the case, a realisation difference would be expected between the older and younger speakers – with the likelihood of mostly younger speakers driving it. This was however not the case, as the outcomes yielded no significance for age, thus suggesting a stability of these phonemes in apparent time.

Though the analysis of the low back vowels, viz. STRUT, LOT & THOUGHT showed a qualitative non-differentiation between LOT & THOUGHT, a gradience of differentiation was revealed between THOUGHT and STRUT. Since this distinction was mainly in F2 – in terms of STRUT movement, the possibility of a unique status was reinforced (see also Jowitt 1991, Eka 2000). As shown in Figure 5.12, the STRUT vowel remains clearly in nearness with the low back classes, and distinct from the low central region (cf. Collins & Mees 1999:2, Fabricius 2007: 296 & Jibril 1986). Such behaviour as suggested in these studies would entail very a high variance in F1, which was not the case in the comparison of STRUT with THOUGHT (Section 5.2.4). Also, despite the significant effect for phone label between the vowels, its value was least ($R^2 =0.257$ p=0.000531), coming behind phone duration and phonological contexts. None of the social predictors was significant in this analysis, thus re-establishing the predominance of linguistic sway in distinguishing between STRUT and THOUGHT.

Concerning the low central clusters, TRAP and BATH were identical, excluding letter which indicated a significant lowering in F1 towards the mid-central region. The lack of contrast between TRAP and BATH is, perhaps, not surprising for EEng, considering reports of similar behaviour in other varieties of English (see also Nycz and Hall-Lew 2014, Hofmann...
In Ugorji’s (2010) proposal reviewed in Section 2.4.2, such fusion is marked for the basilectal accents, but noted as separate phonemes for both mesolectal and acrolectal speakers. Interestingly, Ugorji’s taxonomy finds a parallel in Banjo (1993) whose categorisation of vowels pronunciation in NigE comprises 4 varieties, i.e., Variety I - IV (Section 2.4.1). While Ugorji’s mesolectal cline equals Variety II speakers (consisting of mainly secondary school leavers), acrolectal speakers include mostly tertiary school graduates and other proficient users of the language. Going by these classifications, my speakers would belong to either Variety III or Ugorji’s class of mesolectal/acrolectal speakers – hence, the basis for a similar behaviour. Though TRAP & BATH showed a strong effect of differentiation by phone duration ($R^2 = 0.273$ $p=2.08e-09$), a comparison with Ugorji’s would be hindered due to the absence of detail as to which parameter, i.e., formant values (F1 or F2) or phone duration (ms) specifically differentiates the vowels in his study. Unfortunately, such evidence is beyond the scope of his analysis since it relied primarily on Optimality Theory (OT) rather than acoustic measurements.

The distinction between BATH and letter is least unexpected, given earlier findings in Adive (1989:17) and Carnochan (1970:224) that support a schwa-syllabic system for Ebira geminate vowels (Section 2.3.1), and the recent re-evaluation of schwa realisation by Educated NigE speakers in Jowitt (2014:42). On other hand, the literature has been fairly loud on the obliteration of letter in mostly Southern systems of NigE (Jibril 1986, 1995, Gut 2004, Awonusi 2004, Akinjobi 2009). Considering the delimitation of this study to Ebira system alone, as well as its methodological divergence from a prior appraisal of letter in Educated Yoruba English (e.g. in Akinjobi 2009:93), not only is the ground for comparing both findings problematic, it is also difficult to conclude as to which of the ethnics (Hausa or Yoruba) the status of letter in EEng is pandering.

Also very importantly, although NURSE realisation in the system demonstrates a clear orthographical split between the low back and the mid classes as previously suggested in Schmied (1991a) and Simo Bobda (2000), each sub-set showed a separation from their adjacent classes, i.e., from the low back group of STRUT & THOUGHT, and BATH in the low central, and also from DRESS in the front mid region (cf. Eka 1985, 2004, Odumu 1987, Jowitt 1991, Simo Bobda 2000, Adetugbo 2004). The regressions for determining whether either of the classes of NURSE_back or NURSE_mid deviates from the flanking vowels yielded highly significant values for phone label: NURSE_BATH ($R^2 = 0.60$ $p=8.32e-11$), NURSE_DRESS ($R^2 = 0.52$ $p=3.83e-33$) and NURSE_THOUGHT ($R^2 = 0.33$ $p=0.0102$). Other significant effects
were phonological co-texts and style. The role of age and gender were insignificant, thus contesting the influence of those variables as suggested in Simo Bobda (2000:41).

Even though all predictors were included in the runs for investigating the vowels’ inventory, the corresponding effects of the key social variables were not taken into account. This was in line with RQ1. In what follows, however, the effects of these predictors, excluding phone label and number of L1 were measured in the inventory. By and large, phonological co-texts, phone duration, age, gender and education degree were shown as significant for high vowels, viz.: KIT, FLEECE, USE, GOOSE and FOOT, while co-texts, phone duration and style were generally differentiating for low and mid vowels, viz.: STRUT, THOUGHT, BATH, letter, NURSE(s) and DRESS. With reference to KIT, FLEECE and GOOSE, especially, a similitude of effect for age, gender and phonological environment is found in Mesthrie’s (2010:28) on GOOSE fronting in South African English, and Hofmann’s (2015:359) on the apparent time movement of KIT in Newfoundland variety.

As has been recurrently shown in both layers of analysis, the effect of phone duration was generally predominant. Although the determination of the vowels’ status was considered in reference to the formant values – as opposed to phone duration (cf. Klatt 1976), the quantitative differentiation in the system lends itself to the pedagogical model of Oral English teaching by Nigerian teachers from the outset (Awonusi 1986). Awonsi has recalled that, in the absence of the British instructors, the teachers at the time relied on textbooks for the correct pronunciation, thus making the work of Jones very useful. As cited in Section 2.0, ‘the teachers’ mental conception of the work of Jones, particularly with regard to distinguishing between affected tense and lax vowel distinctions influenced the model that was taught, i.e. it resulted in the interpretation of the tense/ lax vowels distinction as that of duration (thanks to the length marks!) rather than quality’ (Awonusi 1986:557). The trend, therefore, is evidently sustained for oral exercises in Nigerian schools, especially at both Primary and Secondary schools where Oral English is generally included in the curriculum.

Conceptually, the overall gleaning from the results provides some justification for Mufwene’s concept of language evolution and ecology, particularly as it concerns the notion of feature pool (Section 4.1.1). Mufwene conceives variation as the continuing consequences of natural selections from competing alternatives found in idiolects (and dialects) of individuals and communities (Mufwene 2003:146), as such reinforcing the possibility of structural discrepancies between and within group behaviours. Similarly important is his emphasis on the within group variation, especially at the individual speaker’s level. For
example, Mufwene proffers that while variation is typical of all communal languages, there is often regularity in idiolects than in communal languages, including the chances of much greater consistency in idiolectal patterns than for dialects (Mufwene 2003:148). Though my focus on explaining the effects of social predictors was primarily on the communal as well as group behaviours, a fleeting look into the speakers’ behaviour revealed wide ranging idiosyncrasies – as certain trajectories were more individually maintained than collectively for the groups.

An additional correspondence with Mufwene’s view involves the lack of resemblance between EEng inventory and what has earlier been reported for Yoruba (Southern Nigerian English) accent. In spite of his parallels between language and evolution, Mufwene notes the exigency of change, often arising from the ‘restructuring processes’ through the history of a variety. In other words, he considers the possibility of modifications that utterly disconnect a variety from the source structures. By restructuring therefore, he foregrounds the role of ecological triggers in variation, and the resultant prospects of change (Mufwene 2003:147).

As the analysis has shown (see Figure 5.12), a somewhat unique system which panders more to the Educated Hausa than the Educated Yoruba variety is supported. Given the fact that the responsibility of English teaching at the advent of formal education hung mainly on Yoruba teachers who also passed on their accents of English to the earliest crop of educated speakers, I had expected a greater reflection of the Southern behaviour, as opposed to Northern features. In contrast, the majority of the vowel patterns, especially with regard to quality differentiation between high back vowels, e.g., GOOSE & FOOT ($R^2 = 0.703$ $p=0.0144$), and between BATH & letter ($R^2 = 0.42$ $p=2.14e-12$), are a lot more Northern. Excepting the effect of age-group, i.e., the probability of differences between the older and younger speakers, the tendencies suggest a differing system which has, in actual fact, moved on.

Another theoretical instantiation pertains to Schneider’s PCE in NigE. As outlined in Section 4.1.4.1, he believes that ‘English in Nigeria has progressed deeply into phase 3, and has nativised strongly, and is still gaining ground at rapid pace’, while further hinting on a progressive transition to phase 4 (Schneider 2007:212). For a variety at this stage, the PCE framework requires among other features, structural convergence – to the effect that it is hence regarded as X English (Nigerian English) as against ‘English in X (English in Nigeria). The model however appears to rule out the odds of internal and social variation until differentiation in stage 5. Although the idea of ‘convergence’ especially the ‘convergence of educated usages’ has found more espousals in the works of Bamgbose (1995) and Banjo
(1995), the prospect of homogeneity has equally been vehemently opposed in other accounts of NigE (see Schmied 1991a, Jowitt 1991, Simo Bobda 2000, 2007, Udofot 2003, Gut 2004, 2007, Awonusi 2004, Jolayemi 2006, Josiah et al 2012, Ugorji 2015). For instance, while the monophthongal inventory supported in the current study stands in contrast with accounts on other NigE systems, budding proofs of sociolinguistic variation were also reported. This kind of difficulty, perhaps, explains a later clearing up in Schneider (2014:17), in which he notes the deficits of the PCE model as ‘[derivable] from the very nature of [any] model, which [often at best] is an abstraction from reality, not reality “itself,” [thus only a] highlight of specific aspects disregarding others (emphasis mine)’.

6.2 Limitations

This study was limited in a number of aspects, most of which pertain to the basis of analysis, scope and methodology. As stated in RQ1, part of my objective was to investigate the monophthongal inventory of Ebira English vowel system. Accordingly, I intended, from the outset, to ultimately explain the patterns in the light of the speakers’ L1. This goal was particularly inspired by Thomas’s assumption about whether the effects of ATR in most African languages (Section 2.3.1) have substrate influences on African English varieties (Thomas 2011:148). Thus, the corresponding hypotheses would include the odds of identical behaviour in F1 and F2 between tense/lax vowels in EEng and the patterns of ATR/RTR vowels. Towards doing a comparable analysis of both systems therefore, I collected a wordlist data for Ebira L1 from the older speakers. However, vocalic patterns in both systems failed to be comparable on methodological grounds. For instance, the Ebira L1 wordlist tokens used for elicitation accounted for the tonal system in the language – which unfortunately, was not replicable for EEng data. Also, the constituents of tokens from wordlist and reading passage later re-coded as ‘formal style’ differed from the structure of my Ebira ATR tokens. Additionally limiting was the exclusion of younger speakers from the L1 corpus which, in effect, made general assessment across the age groups difficult. Consequently, the EEng system could not be explained in the context of the speakers’ L1, thus complicating empirical grounds for cross-linguistic comparison or the measurement of substrate features from the primary system.

A similar problem concerns the depth of procedural involvement in previous findings. Given the nascence of this study in terms of sociophonetic exploration of NigE vowels,
formulating research questions from previous outings was highly constrained – hence, the expediency of current findings on the strength of previous impressions (e.g. for Yoruba, Hausa and Igbo English accents). Against this backdrop therefore, it became desirable to have included speakers from the ethnics, especially in view of methodological differences between existing literature in NigE and my study. Regrettably, the scale of such prospect was well beyond this study, as only Ebira speakers, thus EEng was investigated. With regard to social grouping, it was difficult reconciling some standard categorisation in variationist paradigms among my speakers. The extent of divergence between standard socio-economic categories in other climes from that of Nigeria has been highlighted above. Likewise, speakers in their post-retirement years could not be included in this study, considering the minimum academic requisite. Their exclusion, however, foreclosed evaluating the trend in which they deviate from other age groups or have loosened up on common behaviours (Labov 1972, Keith 1980). As explained in Section 3.2.1, very few speakers in this cohort would own a tertiary school certificate or might speak English fluently.

6.3 Outlook

As pointed out in 2.6.1, the core objectives of this work include initiating a standard procedure for the investigation of NigE varieties. Though efforts in this area seem relentless, the missing links have been the use of corpus-based data in the variationist assessments. Coupled with a logical re-engagement of this study, some renewed focus is evenly much needed on the inventories of the so-called minor systems of NigE, so as to further understand the wide complexities that constitute the hub of NigE accents.

With respect to social differentiation, it will be worth exploring the variances between speakers who have exclusively been in private schools and those who attended only public schools, as well as the probable effects of urbanity in speech behaviours. For instance, given the diverse mesh of ethnic formations in corporate towers of Nigerian cities such as Lagos, Abuja, Port Hacourt and Kano, it would be desirable to investigate the levelling effects the speakers’ geo-social identity over ethnicity. For a system marked as yet on the brink of endonormative stability (Schneider 2007:212), continuing efforts towards understanding some key social predictors with overriding effects are particularly imperative. This way, further analysis of varieties in the system can be done in targeted perspective rather than prior throngs of bland quests for unifying features.
In further efforts to understand the cross-linguistic interactions between EEng and the speakers’ L1 system (as proposed in Thomas (2011:148)), the need to re-engage the measurement of tense/lax vowels alongside their corresponding ATR/RTR vowels in the primary system remains crucial. For example, an additional investigation of substrate effects of speakers’ L1 vowels on the L2 (English variety) would be highly complementing to this study as such enquiry would either substantially confirm or refute the perennial claims of mother-tongue interference in second language situations, especially in NigE. Ultimately, it would be very fulfilling to stretch the scope of the variationist framework as deployed in this study to include the prosodic layers of the NigE. Though the study of lectal variation has mostly focused on the segmental properties of speech behaviours – with only scrubby emphasis on suprasegmental aspects such as stress, intonation and rhythm (cf. Gut 2001, Jolayemi 2006), a further sociophonetic review of variables assessed in this study in future works would be timely and research-worthy.
Bibliography


Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. Language Variation and Change. 2/2, 205–254.


Taiwo, Rotimi. 2009. The functions of English in Nigeria from the earliest times to the present day. English Today 98, 5/2. 3-10.


Deutsche Zusammenfassung

1. Thema


FF1) Welches Vokalinventar besitzt Ebíra Englisch?


Auf der Grundlage von fast 15.000 extrahierten Vokalen erfolgte jeweils nach der sorgfältigen Aussortierung unbrauchbarer oder unvollständiger Daten eine quantitative Untersuchung mit Hilfe des Analyseinstruments PRAAT, mit dem sich die Vokalqualität in Form von Formanten messen und darstellen lässt. Die Untersuchung umfasste die bekannten Monophongkontraste (nach Wells’ lexical sets) FLEECE & KIT, FOOT & GOOSE (+USE), LOT
Welche sprachlichen und sozialen Variablen können die Variation dieses Ebíra Englisch Vokalsystems erklären?

Neben den bekannten sozialen Variablen Alter (bzw. Altersgruppe), Geschlecht, Mehrsprachigkeit und Bildung wurden v.a. die sprachlichen Variablen Vokaldauer, phonetische Umgebung der Vokale und Sprachstil untersucht. Interessanterweise war für eine so detaillierte Analyse der Variation die zunächst recht groß wirkende Anzahl der extrahierten Vokale nicht in jedem Fall groß genug oder nicht gut genug verteilt.

2. Aufbau

werden könnte, z.B. durch eine stärkere Einbeziehung verschiedener Erstsprachen oder prosodischer Aspekte von Wortbetonung bis Satzrhythmus.

3. Ergebnisse


Interessanterweise brachte die Variable Alter (auch aufgrund der Verteilung der Informanten) relativ wenige Ergebnisse, v.a. weil die stärkere Betonung des Englischen im Unterricht in Ebiraland erst Mitte der 90er Jahre erfolgte und deshalb leider nur wenig Einfluss selbst auf die jüngsten Informanten haben konnte. Nach diesen Ergebnissen müsste eine relative Stabilität der Vokale des nigerianischen Englisch angenommen werden, was vielen anderen Studien aber auch verschiedenen theoretischen Konzepten des postkolonialen Englisch widerspricht (z.B. dem Dynamic Model).

Eines der wenigen klaren Ergebnisse dieser Untersuchung ist der mangelnde Kontrast zwischen TRAP und BATH, was allerdings von vielen anderen vergleichbaren empirischen Arbeiten bekannt ist. Bei einigen anderen Vokalen bzw. Vokalpaaren sind ähnliche Tendenzen festzustellen.

Prinzipiell erbrachten die sprachlichen Variablen Vokaldauer, phonetisches Umfeld und Sprachstil klar signifikantere Resultate als die sozialen Variablen. Dies ist für die Erforschung des nigerianischen Englisch allgemein von besonderer Bedeutung, weil diese Dimensionen bisher in den Untersuchungen meist vernachlässigt wurden.

Auf Grund des Umfangs der Studie, wird ein großer Teil der weniger zentralen Forschungsdetails im Anhang aufgeführt, wie zum Beispiel die vollständige Erläuterung der Datenaufnahme (vom erklärenden Einführungstext bis zur abschließenden Einverständniserklärung der Informanten), die „Entscheidungsbäume“ für die Ausprägung der Vokale nach der Wichtigkeit der Variablen, weitere Statistiken von Regressionenanalysen und Abbildungen von Vokalunterschieden zwischen den einzelnen Gruppen.

4. Ausblick

CURRICULUM VITAE

PERSONAL DATA

Name: ISIKA, Lasisi Adeiza
Date of Birth: 14th December 1983,
Place of Birth: Ado-Ekiti, Ekiti State
Nationality: Nigerian
Local Government: Okene Local Government, Kogi State
Address: English Language and Linguistics, D 09107
Chemnitz Reichenhainer Str. 39, Zi. 208
Email: lasisi-adeiza.isiaka@s2013.tu-chemnitz.de
Alternative e-mail: adeiza.isiaka@yahoo.com

EDUCATION

Adekunle Ajasin University, Akungba Akoko 2004- 2008
University of Lagos, Akoka 2010 – 2011
Chemnitz University of Technology 2013 – Date
DAAD/InProTUC Exchange Programme
Tulane University, New Orleans, USA October – November 2017

Academic Qualifications Obtained

B.A. Ed. (Hons.) English Studies Education
(Akungba), Second Class Upper 2004- 2008
M.A. English (Language) (Lagos) (PhD
Grade 4.28 on the Scale of 5.00) 2010 - 2011

Awards/ Honours

Top Seven Paper Award, ICA 63rd Annual Conference, London, 2013
Certificate of Honour for Community Development, NCCF 2009
Certificate of National Service, NYSC (Serial No: A001264145) 2009
Merit Award for Outstanding Corps Member, NYSC 2009
Certificate of Excellence, Microsoft Office, APT Solutions 2008
Best Graduating Student, English Education Department, AAUA 2007
Best Roving Correspondent, Valley Gong Magazine 2004

WORK EXPERIENCE

Extra-curricular work Experience

Facilitator, Language Learning Strategies and Western Teaching Methodology 2010
Head, Language Proficiency Department, D’ SAM, Lagos 2011
Language Specialist, Diction Educator & Literacy Teacher, Chrisland Schools, Lagos 2009
President, Association of Arts Education Students 2007
Roving Correspondent, Valleygong Magazine  
*University Teaching Experience*

Assistant Lecturer, Adekunle Ajasin University  
2011 – 2013

**Courses Taught**

<table>
<thead>
<tr>
<th>CODE</th>
<th>TITLE</th>
<th>UNITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENG 403</td>
<td>Contrastive Analysis</td>
<td>2</td>
</tr>
<tr>
<td>ENG 406</td>
<td>Speech Writing</td>
<td>2</td>
</tr>
<tr>
<td>ENG 303</td>
<td>Applied Linguistics</td>
<td>3</td>
</tr>
<tr>
<td>ENG 305</td>
<td>Sociolinguistics</td>
<td>2</td>
</tr>
<tr>
<td>ENG 308</td>
<td>Language and Communication</td>
<td>2</td>
</tr>
<tr>
<td>ENG 205</td>
<td>History of the English Language</td>
<td>2</td>
</tr>
<tr>
<td>ENG 203</td>
<td>Advanced English Composition</td>
<td>2</td>
</tr>
<tr>
<td>ENG 201</td>
<td>Introduction to Phonetics and Phonology I</td>
<td>3</td>
</tr>
<tr>
<td>ENG 204</td>
<td>Introduction to Phonetics and Phonology II</td>
<td>3</td>
</tr>
<tr>
<td>ENG 101</td>
<td>English Language I</td>
<td>2</td>
</tr>
<tr>
<td>ENG 102</td>
<td>English Languages II</td>
<td>2</td>
</tr>
<tr>
<td>GST 101</td>
<td>Communication in English I</td>
<td>2</td>
</tr>
<tr>
<td>GST 122</td>
<td>Communication in English II</td>
<td>2</td>
</tr>
</tbody>
</table>

**Schedule of Duties Covered:**

i. Taught all assigned courses and tutorials, and conducted tests, including participation in examination supervision.

ii. Performed sundry duties as the Academic Staff Secretary to English Studies Department – which included, taking records of all meeting proceedings and custody of official documents as directed by the HOD.

iii. Coordinated and provided official assistance to students as a Level Adviser.

iv. Conducted research in aspects of Critical Discourse Analysis (CDA) and Sociophonetics.

**Community Service**

Facilitator, UBEC/SUBEB Re-training of Teachers and School Managers on Language Proficiency, Ondo State  
2012

Community Service Coordinator, Nigerian Christian Corpers Fellowship, Yobe State, Damaturu  
2008 – 2009

**MEMBERSHIP OF PROFESSIONAL BODIES**

Member, International Communication Association (ICA), Washington DC, USA
Member, African Urban Youth Language Research Team, University of Cape Town, SA
**PUBLICATIONS**

**Thesis/ Long Essay**


**Published Book(s)**


**Contribution to Book(s)**


**Published Article(s)**


**Papers Accepted for Publication**


SELECTED CONFERENCES


Language Skills

English: competent in reading, writing and speaking
Yoruba: proficient in reading, writing and translating
Ebira: highly proficient in speaking, reading and writing

Chemnitz, den _______________ Signature _______________

Chemnitz, den________________                          Unterschrift__________________