LANGUAGE VARIATION IN SOUTH AFRICA:
A SOCIOPHONETIC STUDY OF THE VOWEL SYSTEM OF BLACK SOUTH AFRICAN ENGLISH

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der
Fakultät für Sprach-, Literatur- und Kulturwissenschaften
der Universität Regensburg

vorgelegt von
Cornelia Neubert
aus
Chemnitz

im Jahr
2020

Erstgutachter:	Prof. Dr. Edgar Schneider
Zweitgutachter:	Prof. Dr. Thorsten Brato

…to Uli

Table of Contents

Tables

Figures

Abbreviations and symbols

/ /	phonemic transcription
[ ]	phonetic/allophonic transcription
< >	grapheme
~	or
ː	long vowel
ˑ	half-long vowel
[ë]	centralised realisation of a vowel
[e̞]	lowered realisation of a vowel
[e̝]	raised realisation of a vowel
[l̩]	syllabic realisation of a consonant
AfkE	Afrikaans English
BATH	example of lexical set (small capitals)
BSAE	Black South African English
CFE	Cape Flats English
F1	formant 1
F2	formant 2
IPA	International Phonetic Alphabet
IN	Interview
ISAE	Indian South African English
L1	first language, mother tongue
L2	second language
LU	Lobanov Unit (the distance between two vowels in a Lobanov-normalised vowel space)
PCE	Postcolonial English
ReP	Reading Passage
RP	Received Pronunciation
SAE	South African English
WL	Wordlist
WSAE	White South African English

Acknowledgements

There are a number of people I would like to thank for supporting me and helping me to complete my dissertation project. First and foremost, I would like to express my sincere gratitude to my supervisor Edgar Schneider for his invaluable feedback and suggestions. I feel honoured to have been his PhD student. I am further deeply indebted to Thorsten Brato, who sparked my interest for sociophonetics and who patiently guided me in all technical, phonetic and statistical matters where I needed help. I also want to thank Josef Schmied, who supported my research project and paved the way for my fieldwork in South Africa. I am also grateful to Susan Coetzee-Van Rooy and Bertus Van Rooy, who warmly welcomed me in Vanderbijlpark, South Africa, and helped me to make my fieldwork a success. I also want to thank all the participants who kindly took part in my study. Without them, this work would not have come about. Furthermore, I am grateful to Karen Glaser and Wolfram Dötzel for their stimulating comments on earlier drafts of this work from various academic perspectives. I want to thank Tracy Rammler for proofreading the manuscript; any remaining errors are solely mine. Thanks are also due to Alexey Shaporin for his invaluable insights into the many wonders of word processing software. Finally, I would like to thank my family for their support, encouragement and love, most of all my dear husband Uli, whom I owe so much. Unfortunately, he did not live to see the end of this project. I dedicate this work to him.

1 "Eish, but is it English?"

… is a question humorously asked (and answered) by Mesthrie & Hromnik (2011) in their book by the same title that celebrates the English varieties of South Africa. Indeed, there is much to celebrate: Of the eleven official languages of South Africa, English is the most important one. It is the country's general *lingua franca* and spoken by many South Africans as a second language. South African English (SAE) comprises no less than five major ethnolects labelled according to the linguistic or ethnic group who speaks it: White South African English (WSAE), Afrikaans English (AfkE), Indian South African English (ISAE), Cape Flats English (CFE, also called Coloured English)[1] and Black South African English (BSAE). WSAE refers to native language varieties of English spoken by white South Africans of British descent. AfkE is a second language variety spoken by white South Africans whose mother tongue is Afrikaans. ISAE refers to the native language as well as the second language variety spoken by South Africans of Indian descent. CFE is the second language variety mainly spoken by South Africans of mixed racial ancestry whose mother tongue is predominantly Afrikaans. BSAE, which is the subject of the present work, is the second language variety of speakers whose mother tongue is one of the Bantu languages indigenous to South Africa. Black South Africans constitute the largest ethnic group in the country making BSAE "the most widely used form of English in contemporary South Africa" (Bekker & Van Rooy 2015: 290) and also a stimulating field of research.

In fact, BSAE is the best researched variety of the non-ancestral South African English dialects (Bekker & Van Rooy 2015: 286) and has received attention on various linguistic levels, such as morphosyntax (e.g. Minow 2010; Siebers 2012), pragmatics (e.g. Kamwangamalu 2001; Makalela 2004), status, attitude and identity construction (e.g. De Kadt 1993; Coetzee-Van Rooy & Van Rooy 2005; Schneider 2007; Coetzee-Van Rooy 2014), and phonetics and phonology (e.g. Hundleby 1963; Van Rooy & Van Huyssteen 2000; Wissing 2002; Brink & Botha 2005; Da Silva 2007; Mesthrie 2009b, 2017). Of all features, however, it is pronunciation, and in particular the pronunciation of vowels, that immediately distinguishes one accent from another (e.g. Mesthrie 2004a: 962; Bekker 2008: 5). Therefore, this work investigates the realisation of vowels in speakers of BSAE. It contributes to the knowledge of this South African variety because so far, only two major works have described the BSAE vowel system.

[1] The name refers to the Cape Flats, a large flat area where Cape Town is situated. Under apartheid, it was the 'group area' for the Coloureds, i.e. people of mixed racial ancestry (Malan 1996: 125).

Hundleby provided the first description in 1963, three decades before the end of apartheid. The second account was given by Van Rooy in 2004, drawing on previous studies such as Van Rooy & Van Huyssteen's (2000), i.e. on data collected shortly after the abolishment of apartheid. Based on Hundleby's, Van Rooy & Van Huyssteen's and Van Rooy's findings, BSAE phonology can be summarised as follows: The contrast between tense and lax vowels, characteristic of native varieties of English, is reduced by neutralisation of this contrast and by replacement of central vowels by peripheral ones. Schwa /ə/, the unstressed central vowel in native varieties of English, is replaced by stressed vowels. Some diphthongs are produced as monophthongs. Consonants are generally realised similarly to native varieties, but simplification of consonant clusters and the devoicing of a voiced consonant in syllable-final position are frequent observations. The intonation structure of BSAE also differs from that of L1 English. BSAE has a syllable-timed rhythm, i.e. the syllables occur at regular intervals, resulting from the transfer of prosodic patterns of the Bantu languages (Hundleby 1963; Van Rooy & Van Huyssteen 2000; Van Rooy 2004). In contrast, most native varieties of English are stress-timed, i.e. the stressed syllables tend to recur at regular intervals.

During apartheid, BSAE developed into a relatively homogeneous variety (Van Rooy 2004: 943), but with the end of racial segregation in 1994, the phonological spectrum of BSAE has begun to expand. Therefore, the above description can only provide a rough characterisation of today's BSAE.

English has always been an important resource in black communities. Under the conditions of racial segregation, it provided access to higher education, offered opportunity for social mobility and was regarded by many Blacks as the language of resistance against and liberation from the apartheid regime (De Klerk 1999: 316). However, the majority of black South Africans had little contact with native speakers of English. In black schools, pupils experienced impoverished language learning environments with unsuitable teaching materials and undertrained teachers. The result of this language policy was more often than not poor English language skills (Gough 1996: 54). Since the desegregation of education in the early 1990s, however, an increasing number of Blacks have enjoyed access to formerly Whites-only schools. Better learning conditions and native speaker input have provided the basis for the emergence of an English variety within BSAE with a phonology significantly different from older forms (Gough 1996: 55).

The political changes in South Africa including the increasing importance of English have had a substantial impact on the development and use of English among the black population.

Bekker & Van Rooy (2015: 291) therefore conclude that “[…] in the present generation, significant changes are likely to occur.” The present study documents the phonological shape of modern BSAE which are assumed to have arisen from the political changes in the country. It employs an acoustic-phonetic analysis of vowel quality and vowel duration of stressed monophthongs

1. to determine the number of distinct vowels of contemporary BSAE
2. to determine durational differences between vowels classified as tense and lax vowels and
3. to investigate the possible influence of social and linguistic factors on the variation in the BSAE monophthong system.

The basis for this study is apparent-time data of the pronunciation of three age groups labelled young, middle and older. The age of the participants can be related to the political system in which they grew up and went to school. Data collection was carried out in 2013, nineteen years after the first free election. That means that at the time of fieldwork, the participants of the young age group had finished school under the new political order. For them, inter-racial interaction and conversation with native speakers of English should have become routine. By contrast, all middle-aged and older participants had consciously experienced a segregated society; some participants even for decades. Against this historical background, the comparison of speakers of different age groups is assumed to elicit differences in their language behaviour in that young speakers are expected to have a larger vowel inventory than the other two groups and in that the middle and older group also differ in this respect.

As mentioned above, this project focusses on monophthongs typically stressed in native Englishes. It leaves the unstressed vowels and diphthongs unconsidered thus running the risk that the description of the BSAE vowel space may remain incomplete. However, this work examines a variety of World Englishes with a variationist sociolinguistic approach by considering apparent time variation. It therefore adds to the knowledge of the phonological properties of BSAE and adopts a research methodology still new in the investigation of African Englishes:

First, it describes an apparent time study of black South Africans. The comparison of speakers of different generations at a single time is an efficient approach to detect language change in progress (Labov 1994: 45–46) and is a new research method in the context of African Englishes. Examples are Fuchs & Gut (2015) and Isiaka (2017) for Nigerian English varieties. Regarding South Africa, previous studies have focussed mainly on young university students (e.g. Van Rooy & Van Huyssteen 2000; Da Silva 2007; Mesthrie 2009b). So far, only Mesthrie et al. (2015) and Chevalier (2019) worked with apparent time data in an acoustic analysis of South African English varieties. The present work is thus an innovative contribution to the study of pronunciation in apparent time in an African English variety.

Second, the speakers were tested with data collected on the basis of a classic Labovian approach, i.e. a sociolinguistic interview consisting of three parts: the reading of a reading passage, the reading of a wordlist and free conversation. This method has been only rarely applied systematically to language research in African Englishes (notable exceptions are for example Isiaka (2017) and Hoffmann (2011)) and can therefore be considered a novelty in this respect.

Third, in an acoustic analysis of speech, it is meaningful to use vowel normalisation techniques. Such techniques compensate frequency differences resulting from the individual anatomical conditions in the vowel tract of the speakers and thus make their speech comparable. Previous works relied on auditory investigations (e.g. Hundleby 1963; Da Silva 2007), worked with unnormalised data (Van Rooy & Van Huyssteen 2000: e.g.) or examined normalised data of single vowels (e.g. Mesthrie 2009b; Mesthrie et al. 2015). This is the first study that used normalised data for the analysis of the BSAE monophthong system.

The following chapter starts with an overview of the evolution of BSAE and provides a detailed description of the BSAE vowel system along with an introduction of important sociophonetic studies in this field. It also includes information about the attitude of South Africans towards BSAE and discusses the position of BSAE within Schneider’s (2007) Dynamic Model of the evolution of Postcolonial Englishes.

Chapter 3 examines parts of South Africa’s linguistic landscape. One focus is on the phonological description of languages of the Bantu language family as the substrates on which BSAE is based. A second focus is on White South African English, the possible norm BSAE speakers may want to achieve (Mesthrie et al. 2015: 4).

Chapter 4 outlines the study design. It explains the composition of the informants and the linguistic and social variables under scrutiny. It furthermore describes the phonetic analysis and statistical models. This chapter also formulates the specific research hypotheses that were constructed against the background of the recent history of South Africa along with sociolinguistic considerations.

Chapter 5 presents the results of the study. The first part deals with the analysis of the questionnaire revealing language use and language attitude of the participants. After that, the

BSAE vowel space is explained. Each section examines an individual vowel cluster and has generally the same structure: It starts with a descriptive analysis of the data and continues with the results from the statistical modelling. Vowel quality is investigated first, followed by the analysis of vowel duration. In some cases, special features of individual vowels are additionally investigated. Moreover, it also presents observations outside the research focus. At the end of this chapter, the validity of the research hypotheses is discussed.

Finally, chapter 6 summarises the major findings, looks at the limitations of the study and offers an outlook for further investigation.

Eish is an interjection of Bantu origin, which has become a pan-ethnic expression. It was often used during the interview sessions by the young participants articulating excitement or surprise. Excitement and surprise were also my companions during the fieldwork and the data analysis, and hopefully they accompany the reader through this work.

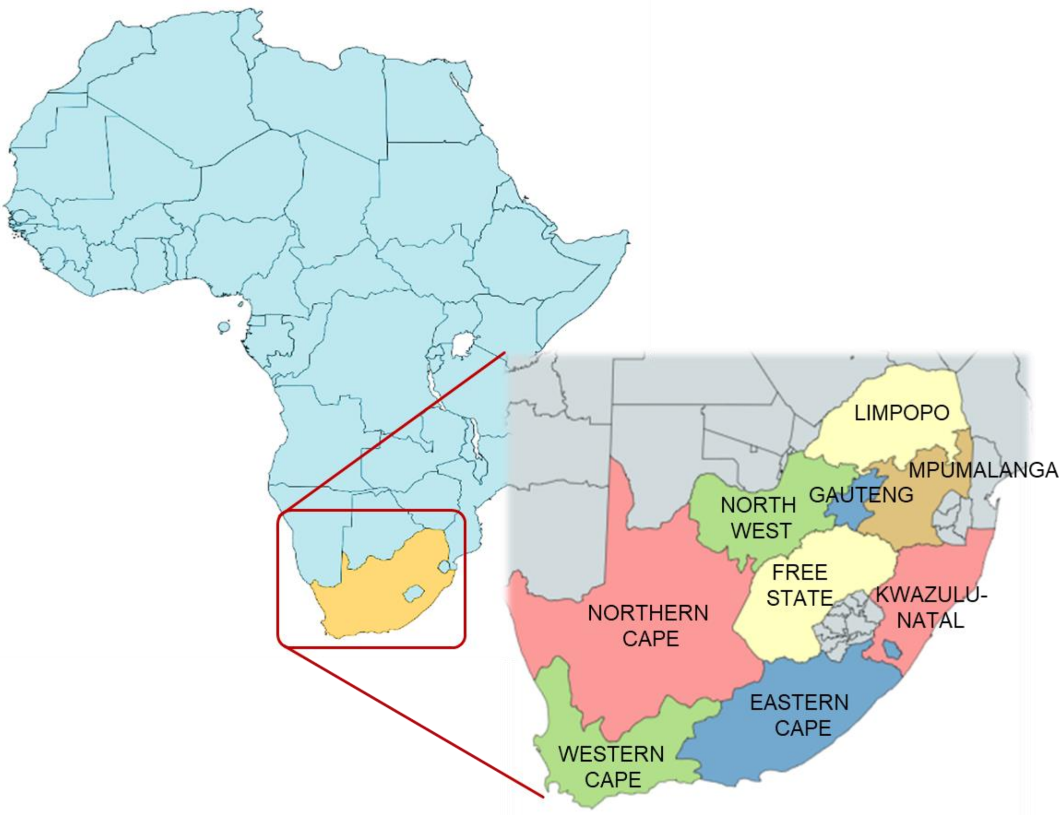

Figure 1.1: Map of South Africa

2 Black South African English

2.1 The environment of acquisition

The evolution of BSAE began at mission schools in the early 19th century (Hirson 1981: 219). The aim of the missionaries was to convert local people to Christianity and introduce "Christian and Western concepts via education" (Mesthrie 2009a: 282). Owing to interaction with well-educated adult and children English native speakers, black pupils had the opportunity to acquire native-like competence (Bekker & Van Rooy 2015: 290). However, due to the limited resources of the church, the missionaries concentrated their educational initiatives within major institutions. By contrast, village schools saw minimal provision. In addition, the colonial governments provided little financial aid for the education of Blacks. Therefore, the number of Africans learning English under appropriate conditions was relatively small. Most black schools experienced constant financial shortage and overcrowding (Hartshorne 1992: 24; De Klerk 1999: 312).

Education of black children was guided by white economic and political interests. The discovery of diamonds and gold in the 1870s boosted a rapid development of urban centres in the mining areas and led to a shortage of labour. In order to meet the demands of the ever-growing mining industry, black pupils had to develop "appropriate work attitudes such as diligence and punctuality" and had to recognise "the operation of the colour-caste system, and their subordinate position … in South Africa" (Hartshorne 1992: 31).

At the turn of the twentieth century, black schools in towns used English as the medium of instruction whereas in village schools, English but also very often an African language was used. The use of Dutch/Afrikaans, the language of Dutch settlers and their descendants, had not been on the agenda yet. The domination of English in black schools was part of the British colonial policy to anglicise the colonies in order to maintain political and economic power (Hartshorne 1992: 189–191). The settlers of Dutch origin, however, struggled for the maintenance and defence of their mother tongue. With the Union of South Africa in 1910, English and Dutch (from 1925 onwards Afrikaans) became the two official languages in the country with equal rights (De Kadt 1993: 312; Hartshorne 1992: 191). This decision had far-reaching implications on black schooling although not immediately. Still, over decades, English was taught as a subject and used as the medium of instruction in all black schools. Afrikaners, however, "had a clear appreciation of the importance of their language as the outward symbol of identity and solidarity. […], and the struggle for Afrikaans became part of the mission to control and rule South Africa" (Hartshorne 1992: 195).

Afrikaner nationalism led to the victory of the National Party in 1948. Apartheid became the governing policy, which also affected education. In 1953, the Bantu Education Act was passed, implementing separate educational strategies and different degrees of financial support for Whites and non-Whites. It stipulated the extension of mother-tongue education from grade 4 to grade 8. It also introduced a strict 50-50 use of English and Afrikaans as media of instruction in secondary schools to ensure equity between these two languages. Both measures had the aim to promote Afrikaans and to reduce the use of English. Dual medium education was not required in white schools though (Hartshorne 1992: 198). The Act was intended to promote white interests on the one hand and to create a black "semi-literate isolated labour force" on the other (De Klerk 1999: 312).

With the policies of the so-called Bantu education, the influence of mission schools declined, and mother-tongue English teachers at black schools became scarce. Most remaining teachers were second-language speakers with limited teacher training who had passed through the very system as their pupils. They often lacked English proficiency on all levels and were unable to act as linguistic models (Hundleby 1963: 144; Wright 1996: 150; De Klerk 1999: 312; Mesthrie & Hromnik 2011: 29). Van Rooy (2020: 228), who gives a summary of the evolution of English and recent trends in sub-Saharan Africa, describes the situation as follows:

> Since English was given such a prominent role in education, the […] transplantation of English […] by African teachers, themselves not always fully proficient in standard English, […] led to the more typical African Englishes that are characterized by a more extensive transfer of linguistic properties of the indigenous languages, as well as all kinds of simplification and regularization phenomena.

In 1975, the government made two major changes in school policy with a tragic outcome. First, it reduced primary education from eight to seven years. Second, mother-tongue instruction was reduced from eight to six years of primary education. From the seventh year (Standard 5), English and Afrikaans were to be used as media in equal parts (Hartshorne 1992: 202). Primary school-leaving certificate examination, which governed admission to secondary education, now took place after Standard 5. Formerly written in the mother-tongue, it now had to be written in English and Afrikaans after only one year of instruction (Hartshorne 1992: 75, 202). Owing to an inadequate preparation of the language switch, pupils had tremendous difficulties with instructions in the second language. The results were high failure rates and high drop-out rates (Kamwangamalu 2004: 226).

Black students saw themselves deprived of the chance to access higher education and thus of social and economic participation. Anger and frustration among the population culminated in the Soweto uprising starting 16 June 1976. It marked the end of Afrikaans as a medium of instruction and boosted the use of English in black schools as well as in black communities (Kamwangamalu 2004: 230; De Klerk 1999: 312; McArthur 2002: 289; Hartshorne 1992: 203).

In 1990, negotiations between the government and the liberation movement African National Congress (ANC) paved the way for a new political order (Mesthrie et al. 2000: 411–412). The ANC soon recognised that language planning was a pivotal question in post-apartheid South Africa. The ANC had initially regarded English as the language of national unity, but during the late 1980s, its language policy had turned to the promotion of the country's indigenous languages. During the apartheid era, English and Afrikaans were the two official languages in the country. Indigenous African languages had regional co-official status in the so-called homelands.

Due to the diverse linguistic landscape of the country, multilingualism was wide-spread but only in terms of indigenous African languages. By contrast, proficiency of English was low, and in the beginning of the 1990s, one third of Blacks had no English skills at all. Adopting English as the sole official language would have excluded these people from "political participation, and would have disadvantaged them in the education system and labour market" (Mesthrie et al. 2000: 412). In order to embrace the linguistic diversity of the new "Rainbow nation", the constitution of 1996 therefore recognised the two official and nine co-official languages of the former regime as the new official languages of the country (Mesthrie et al. 2000: 412; Cook 2009: 98).

The South African constitution of 1996 supports a policy of multilingual inclusion and seeks to promote the status of previously disadvantaged languages (Cook 2009: 96). The eleven official languages reflect the major ethnic and language groups in the country, Bantu and Germanic. While the Germanic languages (Afrikaans and English) include 23.1% of speakers, the Bantu languages (Zulu, Xhosa, Ndebele, Swati, Southern Sotho, Northern Sotho, Tswana, Tsonga, Venda) comprise 74.9% of speakers. Linguistically, Afrikaans enjoys a special status among these languages since it is a language of European origin but can be regarded as an indigenous language of Africa (Mesthrie 2002: 5–6; McArthur 2002: 266).

In post-apartheid South Africa, a non-racial education system has been established: The schools are open to all races, and multilingualism is promoted. From grade 3 onwards, pupils study the language of instruction plus at least one other language (Kamwangamalu 2004: 231). The practical implementation of this well-intentioned agenda, however, caused many problems: The curricula promoted multilingual skills at school where single-medium education had been the norm until then. Hence, adequately prepared teachers and suitable teaching material were rare. Therefore, black parents strove to send their children to formerly white schools where they were taught in English from an early age and received a good education (Kamwangamalu 2004: 233). Apart from that, even in predominantly black schools, African languages as the media of instruction in primary schools have been increasingly replaced by English, a development that is greeted by many parents. In their opinion, good English skills are the key to social mobility and economic progress (Mesthrie et al. 2000: 415; Kamwangamalu 2004: 233, 2009: 163–164). By contrast, education in African languages is not considered beneficial for success in life (Bamgboṣe 2009: 652). This attitude has led to a decrease in competence in the mother tongues (De Klerk 1999: 319).

Despite the achievements in the education of non-white people, not every child can profit from good English learning conditions, the consequence of which is a continuum of second-language English varieties ranging from basic or functional forms of English to L1 qualities. Apart from that, most adult Blacks experienced a segregated school system. Therefore, "[t]he long-term effects of under-funding, overcrowding and teacher incompetence, combined with limited contact with native English speakers, led to characteristic patterns of pronunciation and syntax becoming entrenched as norms of spoken BSAE, with resultant reduced levels of comprehensibility" (De Klerk 1999: 312; Wright 1996: 151).

Concerning the numbers of BSAE speakers, it is difficult to provide exact figures. The last national census identified 1.17 million black L1 speakers of English, which is 2.9% of the 40.4 million black South Africans (Census 2011: 25-26). However, there are no current statistics on the proportion of black South Africans who have a knowledge of English. Former figures range from 32% in the 1991 national census to 61% in the 1993 RCM survey (Reaching Critical Mass Survey Report cited in Gough 1996: 53). Regarding spoken competence in first- and second-language English, De Kadt (1993: 314) assumes 29% of the whole South African population. Thus, this figure must have been much lower for the black population. What is undisputable though is the fact that the use of English as an additional language is rapidly increasing (De Klerk 1999: 314).

2.2 Attitudes towards BSAE

English varieties in South Africa are divided along ethnic lines, such as White, Black or Indian. These labels seem reasonable since the varieties they represent exhibit distinct phonological features. On the part of BSAE, for example, Van Rooy (2004: 943) states that "segregative and oppressive practices of apartheid had led to the development of a relatively homogeneous second language variety", but he also points out that today it "is becoming more diffuse" (Van Rooy 2004: 943). Up to the 1990s, BSAE had been considered as an English variety that deviates from the norm and is unacceptable in formal contexts, particularly because of its phonology that differs greatly from the normative WSAE (De Klerk & Gough 2002: 356). After the abolition of apartheid, the view on BSAE changed, and it has been since largely recognised as a self-sufficient variety (Gough 1996: 70; Mesthrie 2010a: 188). The use of a non-standard variety, which was stigmatised in the past, has become more and more accepted and acknowledged as part of the South African English system. Factors such as the use of English as a major language of government (where L1 English speakers are rare), its increasing use in official announcements in the media, and the emergence and growth of a black middle-class have contributed to this development. As a result, a prescriptive attitude has been in parts replaced by a more tolerant mindset. The changing perception of BSAE in public life such as media and business has led to an increased use of BSAE accents as well (De Klerk & Gough 2002: 370; De Klerk 1999: 317).

However, there are still conservative voices which strive for the maintenance of prevailing language standards and deny the viability of BSAE. Another counter movement is expected from those of the black youth who attend former white English-medium schools. By the time they leave school, they will have acquired a standard SAE variety and will probably support a normative (maybe even an exonormative) English rather than an "ethnically marked BSAE" (De Klerk & Gough 2002: 371). Finally, many less privileged L2 English learners may regard English only as an instrument necessary for competition on the job market and not as a means for integration in society (De Klerk & Gough 2002: 371).

Attitudes towards BSAE have been the subject of both qualitative and quantitative studies. In 1993, Smit (1996) investigated the attitude of 282 high school pupils of different ethnicity towards BSAE, AfkE and L1 English. L1 English scored highest in the categories that were associated with education and socio-economic achievement. The question of whether English should take words, meaning and structures from African languages was accepted and rejected by half of the respondents (50.4% 'yes' to 49.6% 'no') respectively. A surprising 65% of white

L1 pupils who attended non-private schools opted for 'yes'. The question of which kind(s) of English should be used in education was answered by 63.5% with "Standard South African English". This variety was preferred in all language groups. An "Africanised English" was largely disregarded, with only 8.5% responding with 'yes'-answers. Smit (1996: 98) concludes that "English still holds a pragmatically determined privileged and central position as main language of national communication and medium of higher education", but she also observed a willingness to support non-standard varieties.

BSAE as a second language variety can be divided into three sub-varieties labelled acrolect, mesolect and basilect, which are regions on a second language continuum (cf. Schmied 1991: 47-49 for African Englishes). Concerning pronunciation, acrolectal speakers perform a vowel contrast similar to that of the particular language norm whereas the mesolect has a smaller vowel repertoire and therefore deviates more from the target language. The basilect represents the lowest command of English (Van Rooy 2004: 944, 2020: 222). Van Rooy et al. (2000) studied the attitude of 106 Sotho, English and Afrikaans speaking high school students towards acro-, meso- and basilectal BSAE. By means of a modified matched-guise test, they found that the English and Afrikaans speaking pupils showed the most positive attitude towards WSAE, but also acknowledged acrolectal BSAE. The Sotho pupils favoured acrolectal BSAE with WSAE coming second. The reason why black pupils prefer acrolectal BSAE is its function as a marker of identity (Van Rooy et al. 2000: 205-206). Moreover, an acrolect is phonemically sufficiently close to the standard variety of WSAE and thus comprehensible. The authors conclude that "[as] BSAE variety of the New South African elite, it might increasingly come to serve as an idealised target for language acquisition by BSAE learners in secondary education […]" (Van Rooy et al. 2000: 206). Mesolectal BSAE received some support by the participants. The reason for this may lie in the evolvement of an urban black middle class whose variety is primarily mesolectal (cf. Da Silva 2007). As this variety may have gained in-group status, it has probably become the linguistic norm (Van Rooy et al. 2000: 206). Basilectal BSAE was unacceptable for all participant groups. Its features convey notions of limited access to education and are associated with the apartheid regime. The authors overall conclude that the status of educated BSAE is increasing. Furthermore, it is greeted even by Whites "as an alternative, or parallel, standard of English in South Africa" (Van Rooy et al. 2000: 208).

Coetzee-Van Rooy & Van Rooy (2005) conducted a study with 167 black Bachelor students of non-linguistic subjects to elicit the labelling of and attitude towards six SAE varieties. Among other things, they wanted to find out whether the label 'Black South African English' is accepted among Blacks. This label is criticised by some linguists because it emphasises ethnicity. A text was read in acrolectal BSAE, mesolectal BSAE, General SAE, Broad SAE, acrolectal Indian SAE and acrolectal Afrikaans English. The respondents were asked to assign their own terminology to the particular SAE variety. As for the naming of acrolectal and mesolectal BSAE, the vast majority used labels that included 'South African', 'African', 'Black', 'Zulu' or 'Sotho' English with 'South African English' receiving the highest number of responses (34 for acrolectal and 43 for mesolectal BSAE). They conclude that the label 'Black South African English' does not provoke resentment among black people and that they use labels that emphasise ethnicity. Interestingly, the most popular label across all six varieties in total was 'South African English', an option that was most frequently used for the BSAE varieties too. The participants also had to indicate their perception of the proximity of their variety of English to the English of the six SAE varieties. Acrolectal BSAE received the highest score although most participants spoke mesolectal BSAE. The authors ascribe this mismatch to the participants' perceived status in their community.

Another outcome of this study was that the participants very often were not able to identify sub-varieties of BSAE. For example, the acrolectal BSAE reader was a Xhosa but was three times identified as Sotho or Tswana. This led to the conclusion that BSAE – although their speakers may have L1s that are not mutually intelligible – cannot be distinguished by their English accents. Finally, in an attitude and comprehensibility test, the acrolectal BSAE enjoyed the highest prestige. Mesolectal BSAE was the least favourable, suggesting that the respondents rated their own accent low (Coetzee-Van Rooy & Van Rooy 2005: 14).

2.3 Embedding BSAE in the Dynamic Model

Within the paradigm of World Englishes, South African English (SAE) can be categorised as a postcolonial variety. As such, SAE, including all sub-varieties, can be described by the Dynamic Model of the evolution of Postcolonial Englishes (PCE) (Schneider 2007). This model proposes a universal pattern that every PCE variety has to undergo irrespective of local idiosyncrasies of the particular case. This pattern includes the following five consecutive phases:

1. Foundation: English is brought to a new territory. The contact between different English dialects on foreign ground leads to koinéisation, the levelling of dialects. The newcomers and the indigenous population regard themselves as separate (Schneider 2014: 11; Bekker 2008: 84).

2. Exonormative stabilisation: The colony is politically stabilised. The contact between settlers and the indigenous people increases. The linguistic norm is still British but flavoured with lexical loans necessary to denote the local environment ("British plus") (Schneider 2007: 11). A small proportion of the indigenous population becomes bilingual (Schneider 2014: 11; Bekker 2008: 84–85).
3. Nativisation: In this phase, cultural and linguistic transformation takes place. The colony gradually moves towards political and cultural independence. The expanded contact reduces the social and linguistic gap between the settlers (and their descendants) and the natives and leads to an emerging variety with innovations in lexis, phonology and grammar (Schneider 2007: 183, 2014: 11; Bekker 2008: 85).
4. Endonormative stabilisation: This stage is reached after political independence and is associated with nation building. The transition from phase 3 to phase 4 presupposes an "Event X", an "exceptional, quasi-catastrophic political event" that leads to the detachment from the mother country and a redefinition of identity that embraces all population groups (Schneider 2007: 48, 2014: 12). This identity surfaces as a new local variety which "is perceived as remarkably homogeneous" (Schneider 2007: 51). Codification of the new norm is manifested by the production of local dictionaries (Schneider 2014: 12).
5. Differentiation: Political stability and internal cohesiveness of the new nation lead to internal differentiation and thus to the emergence of new dialects and sociolects (Schneider 2014: 12; Bekker 2008: 85).

The key players in these scenarios are two groups – the colonisers (the settlers, "STL") and the colonised (the indigenous residents, "IDG"). Both groups experience the development of English from two complementary perspectives (Schneider 2007: 31). In the case of South Africa, Schneider points out that the sociolinguistic constellation in the country is extremely complex. He explains:

> This complexity arises not primarily from the country's multilingualism but rather from the fact that a comparatively high number of distinct, compartmentalized speech communities have entered the arena at different points in time and have interacted with each other under varying social circumstances (Schneider 2007: 174).

The following will give a brief summary of the stages and concentrate on the current sociolinguistic situation in South Africa. A thorough description of the historical development of South African English is offered by Schneider (2007).

The history of English in South Africa (Foundation phase) began with the final British seizure of the Cape colony in 1806 (Lanham 1996: 19; McArthur 2002: 288; Schneider 2007: 175; Bekker 2013). The first European settlers in the Cape region, however, were the Dutch, who had arrived in 1652. When the British came to South Africa, they thus encountered two different resident groups indigenous to the territory: African tribes and the Afrikaners (or Boers), the descendants of Dutch immigrants, who had lived there for generations (Schneider 2007: 176). The contact with the Dutch/Afrikaners and the Africans resulted in the "Coloureds", a distinct ethnic group of mixed European-African ancestry. Their mother tongue was mainly Afrikaans, but they also spoke one or more Bantu languages. With the arrival of the British, this group also learnt English and thus contributed to bi- and multilingualism (Schneider 2007: 177).

The largest group of the IDG strand were the African tribes. Ever since the Europeans set foot on South African territory, they were confronted with the presence and encroachment of European settlers. As a consequence, there were numerous frontier wars during the eighteenth and nineteenth century. During that time, contact with the settlers were scarce, and the Africans largely continued with their traditional lifestyle (Schneider 2007: 176).

The first systematic immigration of British subjects to South Africa took place in 1820. People of largely lower-class descent from southern England settled in the Eastern Cape. The proclamation of 1822 to implement English as the only official language in the colony started the phase of exonormative stabilisation (Schneider 2007: 178). The language of the descendants of the 1820 settlers became the first local native variety of English in South Africa (Lanham 1996: 20–21). The second wave of English immigration took place east of the Cape Colony in Natal between 1848 and 1862. These settlers came from the Midlands and northern counties and were mostly members of the middle and upper class. During this phase, stable colonial dialects with British norms were emerging (Van Rooy & Terblanche 2010: 359; Schneider 2007: 38-40).

The emerging mining industry during the 1870s led to a rapid development of urban areas and to the fast spread of English. The black population learnt English in the growing number of schools and in various contact situations in urban settings. The Afrikaners, who lived mainly in rural areas in the South African interior, had little share in the prospering economy of the cities. This uneven development gave rise to Afrikaner nationalism and led to the Anglo-Boer Wars between 1899 and 1902, in which the British defeated the Afrikaners (Watermeyer 1996: 102–103; McArthur 2002: 289). The Act of Union of 1910 united the

British colonies and the Boer republics and formed the Union of South Africa (Lanham 1990: 25; Schneider 2007: 178). This event symbolises the onset of phase 3, nativisation, in which the influence of the mother country diminished (Schneider 2007: 178, 2007: 182). During this phase, "[s]tructural nativisation" was powerful on the level of phonology, lexis and grammar and "produced a set of social and ethnic varieties with shared features and some features of their own" (Schneider 2007: 183).

Between 1860 and 1911, another immigrant group came to South Africa. Indentured workers from India were hired to work on sugar plantations in Natal and stayed after their contract had expired. This group were neither English-speaking settlers nor part of the indigenous population, which is why Schneider (2007: 182) calls this strand Adstrate ("ADS"). The Indians maintained their cultural heritage, but quickly learnt English as a means of socioeconomic advancement (Schneider 2007: 176).

Despite the very complex linguistic setup in South Africa, the various speech communities have in common that they experienced a radical political transformation, which was the end of apartheid with the first free election in 1994. This was Event X, the point that introduces the phase of endonormative stabilisation (Schneider 2007: 48, 185). Event X is usually associated with political independence from the mother country. While South Africa had been formally independent since 1961, for the first time in the history of the country, it gave the population a feeling of togetherness and "collective identity" (Schneider 2007: 185). This so-called Rainbow Revolution introduced "a phase of nation-building, cherishing moves towards increasing uniformity, with local educated accents increasingly coming to be accepted in public contexts, codification on the way" (Schneider 2019).

Schneider (2007: 174) locates South African English deeply in phase 4, but he also argues, that due to the different ethnic varieties of SAE, it may accommodate several PCEs. His conclusion is based on the fact that South Africa has no local English norm to express the new identity (Schneider 2007: 185). Schneider's diagnosis coincides with McArthur (2002: 291) who suggests: "There is no South African standard as such for English. Traditionalists look towards the BBC and RP, while American English has had the same kind of influence in the RSA [Republic of South Africa] as elsewhere" (parenthesis added).

The question of whether SAE has reached phase 4 or not has been researched in several studies, mostly based on lexis and morphosyntax. Van Rooy & Terblanche (2010), for example, examined a newspaper corpus. The criterion for the acceptance of particular words in this corpus was their entry in the South African Concise Oxford Dictionary. The authors found considerable evidence for lexical innovations frequently originating from the culture of the indigenous population. They recognised a linguistic convergence between the STL and IDG strands and therefore suggest the emergence of phase 4 for SAE.

Spencer (2010) investigated the acceptability of BSAE features in academic writing among non-black South African university students. The results indicated tolerance towards some features, e.g. the extended use of the progressive aspect or the omission of the article. She argues that even a low level of acceptance of BSAE features in academic works is a sign of linguistic convergence. Spencer therefore sees this variety moving towards phase 4.

The extension of the progressive aspect was also investigated by Kruger & Van Rooy (2017). They compared unedited and edited versions of BSAE texts and found that formerly framed 'errors' are increasingly regarded as innovations. They conclude that the innovative uses of the progressive aspect enjoy "tacit acceptance", which in turn contributes to the achievement of endonormative stabilisation (Kruger & Van Rooy 2017: 38).

Coetzee-Van Rooy (2014) examined the identity construction of university students and found that the home language is the carrier of ethnic and cultural identity. English belongs to the linguistic repertoire of South Africans, but multilingualism (not solely English) is the linguistic marker of a South African identity, an opinion that is also expressed by Schneider (2007: 188). Therefore, she places SAE in phase 3 (nativisation) and sees no indication for South Africa to progress past this stage in the near future (Coetzee-Van Rooy 2014: 54).

Bekker (2008: 86) postulates that the sub-varieties of SAE should be allowed "to run their own 'course'". In this respect, he puts BSAE in phase 3. One characteristic of this phase is that the gap between the STL and the IDG strands is significantly reduced, and Bekker (2008: 86) thus argues that this gap "has *begun* to be bridged" only since the end of apartheid. By contrast, he locates WSAE far into phase 4 (Bekker 2008: 432–433).

Du Plessis (2015) carried out an acoustic-phonetic study of vowels of young white upper-middle class people from Cape Town, Durban and Johannesburg. He documented regional differences between the three major cities and therefore claims that White South African English has reached phase 5 (Du Plessis 2015: 181; see also Du Plessis 2019). Furthermore, he found evidence that males are ahead of females in the shift away from a national norm towards regional dialects (Du Plessis 2015: 179).

Van Rooy (2014) investigated the contact settings of L1 English with other languages spoken in South Africa. He argues that due to the segregation of ethnic groups, Afrikaner nationalism and the strong bonds of the English population to Britain, the rewriting of identity began only in 1994. But with the persistent social tension in the country, a common identity is still symbolic rather than actual practice, and national reconciliation has not been achieved. Yet, in respect of the developments since 1994, Van Rooy (2014: 35) predicts a movement towards endonormativity. At the same time, however, he points out that various contact scenarios have led to the emergence of different stabilised dialects and ethnolects respectively which is why homogeneity will be difficult to achieve, and sub-varieties will not necessarily converge (Van Rooy 2014: 33). Concerning a linguistic norm, White SAE as well as the established ethnolects of Indian and Coloured SAE may develop in parallel. By contrast, BSAE is not stable enough to compete, in particular with WSAE (Van Rooy 2014: 35).

Endonormativity is connected to the tolerance towards linguistic features, most of all to pronunciation (Schneider 2007: 187). As illustrated in Section 2.2, BSAE has gained some support in this respect. However, despite being the biggest ethnolect in the country, it seems that BSAE will not gain the power to develop into a South African norm. The standard will probably be a non-racial variety as "there is a recognized range of usage that fits the label *'educated' South African English*" mainly spoken by "South Africans of largely British descent", but also by "speakers of Afrikaans, Indian languages, and indigenous languages who are comfortably fluent in English" (McArthur 2002: 291; see also Bekker 2008: 88). This "'educated' South African English" is based on L1 WSAE. Non-white speakers move towards greater conformity with the "white values", but corresponding adjustments towards "black values" do not seem to occur (Van Rooy 2014). This development is also reported by Da Silva (2007) and Mesthrie (2009b) and will be discussed further in Section 2.5.2.

2.4 The vowel systems of the substrate languages

Language learners tend to transfer linguistic features of their mother tongue to the foreign language (Lado 1957: 2; Hundleby 1963: 12). This section therefore describes the vowel system of the native languages, which phonologically influences the performance in the target language. BSAE is a second-language variety. The mother tongues of BSAE speakers are indigenous South African languages of the Bantu phylum. The parent group is Central Narrow Bantu S (Bantu Zone S 2014). The major child subgroups spoken in South Africa are

- Sotho-Tswana including the languages Northern Sotho, Southern Sotho, Tswana and Birwa,
- Nguni including the languages Ndebele, Swati, Xhosa and Zulu,
- Tswa-Rhonga including the language Tsonga, Tswa and Ronga, and

- Venda including one language, Venda (Lewis et al. 2014; Gowlett 2006: 609–610).[2]

Dialects of these languages are not known. Northern Sotho and Tswana are inherently intelligible but have been considered separate languages for political reasons (Lewis et al. 2014; Cook 2009: 97). Bantu languages vary in their number of contrastive vowels, but have "a perfectly balanced system with one low vowel, a, and an equal number of equally positioned front and back vowels" (Doke 1967: 26).

The Nguni languages use a five vowel system, /i, e, a, o, u/ (Gowlett 2006: 611; Maddieson 2006: 15). Hundleby (1963: 48) and Doke (1967: 26) describe an /i, e, ɑ, o, u/ system. The phoneme /e/ splits into [ɛ~i] and /o/ into [ɔ~o]. The raised allophones [i, o] are produced when the next syllable contains a close vowel, (/i, u/) (Gowlett 2006: 611). Doke (1967: 27) calls this a system of seven vowels and five phonemes, in which [e, ɛ] are allophones of /e/ and [o, ɔ] are allophones of /o/.

The Sotho-Tswana vowel system is reported to distinguish either seven vowels /i, e, ɛ, a, ɔ, o, u/ (Tucker 1929: 23-26; Arellano 2001: 16; Barnard & Wissing 2008: 255) or nine vowels /i, ɪ, e, ɛ, a, ɔ, o, ʊ, u/ (Clements & Rialland 2008: 17; Doke & Mofokeng 1974: 2; Gowlett 2006: 611). Doke & Mofokeng (1974: 1–2) describe the same system but with the phoneme /ɑ/ instead of /a/. Khabanyane (1991: 8) also describes nine vowels but with different qualities /i u e o ɛ ɛ̹ ɔ ɔ̹ a/. Arellano (2001: 16) argues that a nine vowel representation does not reflect the current system. For Tswana, he distinguishes seven phonemes. Although for him it is acceptable to describe the system with /i, e, ɛ, a, ɔ, o, u/ (cf. Tucker 1929), he suggests the phonemes /i, ɪ, ɛ, a, ɔ, ʊ, u/ because these qualities correspond with the actual pronunciation.

Venda employs five vowels /i, e, a, o, u/. The phoneme /e/ can be represented as [e] or [ɛ], and the phoneme /o/ can be represented as [o] or [ɔ] (Doke 1967: 26, 154). Venda has thus the same vowel system as the Nguni languages.

Southern Bantu languages lack diphthongs and lax vowels as well as central vowels. Vowel length is generally not phonemic, but the penultimate syllable is lengthened due to the phenomenon of length-stress prominence (Van Rooy 2002: 145; Hundleby 1963: 64; Van Rooy & Van Huyssteen 2000: 22; Doke 1967: 26). Bantu languages are syllable-timed, which means that their syllables are similar in length. Figure 2.1 shows the vowel space for Nguni and Sotho-Tswana. The phonemes in brackets belong to the seven-vowel system; Venda is identical with the Nguni system.

[2] Alternative names for some of these languages are Sepedi (Northern Sotho), Sesotho (Southern Sotho), Setswana, isiNdebele, SiSwati, isiXhosa, isiZulu, Xitsonga and Tshivenda (Census 2011: 23).

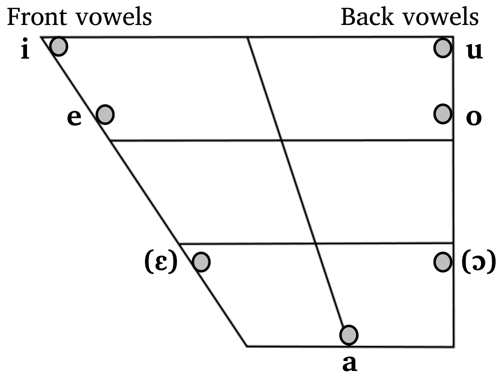

Figure 2.1: The five-vowel system of Nguni and the seven-vowel system of Sotho-Tswana (adapted from Khabanyane 1991: 2)

In contrast to the Bantu system, Standard British English has a vowel set of twelve monophthongs (including central vowels) and eight diphthongs (Jones et al. 2006). It is stress-timed, which means that stressed syllables recur at regular intervals. The difference in the vowel inventory and prosody between both language systems influence L2 English pronunciation.

2.5 Vowel phonology of BSAE

2.5.1 Describing the system

BSAE vowels derive from the transfer of the vowel inventory of the respective first language with similar vowel phones (Van Rooy & Van Huyssteen 2000: 19). Pioneering work on BSAE phonology was carried out by Hundleby (1963). Based on an auditory investigation of nearly 2,000 Xhosa school children and university students in the Eastern Cape, he provided a close phonetic description of Xhosa-English. This study had an educational purpose. Through the analysis of the phonetic and prosodic features of Xhosa-English, Hundleby wanted to provide teaching suggestions for the English classroom, in particular to improve BSAE pronunciation, that is, in his case Xhosa-English pronunciation. He set out with a comparison of the sound system of Xhosa, Received Pronunciation (RP) and White South African English respectively.

He found that

- certain English vowel allophones do not exist in Xhosa,
- certain Xhosa vowel allophones do not exist in English and
- certain English phonemes have an approximate counterpart in Xhosa (Hundleby 1963: 61).

Thus, the differences in the vowel inventory of Xhosa and English lead to the loss of phonemic contrast in Xhosa-English. In order to pronounce an English sound that has no equivalent in Xhosa, the first attempt is to use a Xhosa sound that comes closest to the target, which may lead to the production of incomprehensible English words. In addition, length contrast is not phonemic in Xhosa, a feature which is also transferred to Xhosa-English.

Hundleby (1963: 63–64) describes the vowel features of Xhosa-English as follows. For the representation of the vowels, Wells' (1982a) lexical sets are used. Lexical sets are written in small capitals and comprise the set of words that contains the vowel in question.

FLEECE /iː/, is produced as [iː] in penultimate syllables before a pause (e.g. *believer, repeated* or *I did not steal it*). Elsewhere, it is shortened to [i·] or even [i] and can be lowered to [ɪ]. The word *feel* in the sentences (1) *Feel the box* and (2) *You are well if you feel strong* is thus pronounced [ɪ] in (1) and [iː] in (2). The absence of phonemic contrast in both quality and quantity is also reflected in the misspelling of English words, e.g. *An axe blade is made of still* (Hundleby 1963: 63–64).

KIT /ɪ/, is produced as [i] in non-penultimate utterance position as in *I did not know the time* and as [iː] in penultimate syllables as in *I did it*. Again, this pronunciation pattern in terms of vowel quality, stress and lengthening can lead to confusion. For example, the phrases *A ship's bell* and *A sheep's bell* are pronounced identically, i.e. [ʃiːp]. In cases where word-final <ed> /əd/ in words like *neglected* or *dedicated* is pronounced [ɪd] in RP, it is produced as spelling pronunciation [ɛd] in Xhosa-English (Hundleby 1963: 65–66).

DRESS /ɛ/, is usually produced as [ẹ] and differs slightly from the L1 English phoneme/ɛ/. This realisation, however, does not lead to phonemic confusion (Hundleby 1963: 66–67).

For NURSE /ɜ/, Hundleby states that it is "one of the most difficult of all to acquire, and it exhibits, probably more than any other XEP [Xhosa-English Pronunciation] sound, the greatest degree of variation from any so-called standard pronunciation" (Hundleby 1963: 78; parenthesis added). Owing to the absence of central vowels in Xhosa, /ɜ/ is realised not by a single Xhosa allophone. Its pronunciation may include modifications of the Xhosa vowel phonemes /e/, /ɑ/ and /o/ (Hundleby 1963: 78). The most frequent allophone is a centralised and half-long [ɛ̈·]. Orthographic <or> in words like *word, work* and *world* can induce lip rounding and thus produce [ø·]. Another rounded variant is [œ·] (Hundleby 1963: 79).

TRAP /æ/, is produced as [e], [ɛ] or [a] and thus diverges strongly from the target due to the lack of an equivalent Xhosa sound. A word like *cattle* /kætl/ is pronounced [kɛtl], and *pack* /pæk/ may range from [p^{h}ek] to [p^{h}ak]. When /æ/ is approximated by [e] and [ɛ], word meaning is affected and spelling errors are likely (Hundleby 1963: 68–69).

BATH and START /aː/, are realised as [a] or [ɑ]. Again, due to Xhosa length-stress prominence, the allophone [ɑː] is used in syllables of strong stress in penultimate positions of an utterance. In non-penultimate and unstressed positions, vowel length is reduced, and *part* may be pronounced [paˑt] or [pɑˑt]. This can impair comprehensibility since word pairs like *pat/part*, *back/bark*, *ham/harm* and *battered/bartered* are homophones in Xhosa-English (Hundleby 1963: 70–71).

STRUT /ʌ/, is approximated in XEP with a centralised [ä], and possibly lengthened to [äˑ~äː] after plosives and with strong stress. Words like *cut, cover* and *butter* may thus be understood as *cart, carver* and *barter* (Hundleby 1963: 76–77).

LOT /ɒ/, is produced as [ɔ] or [ɔˑ]. When LOT in penultimate position is followed by /ɪ/ in the following syllable as in *knotty*, *dropping* or *stop it*, the LOT allophone is lengthened to [ɔː] and /ɪ/ becomes a prolonged [iː], rendering the word *knotty naughty*-like. Orthographic <a> in words like *watch* or *want* is often produced as [a] (Hundleby 1963: 71–72).

THOUGHT, FORCE, NORTH /ɔː/, are realised as [ɔː] by educated speakers, and [ɔˑ], [ɔ] or [ɒ] by less educated ones producing [cɔt] for *caught* and [tɒt] for *taught*, for example (Hundleby 1963: 72–73).

GOOSE /uː/, is realised as a lengthened [uː] in penultimate syllables of words and utterance groups. In other positions, it is realised as [u] (Hundleby 1963: 73–74).

FOOT /ʊ/, is also realised as [u] and [uː] under the conditions mentioned for GOOSE. Due to the absence of qualitative and quantitative contrast between the two phonemes, the words *fool* and *full* are homophones and pronounced with [uˑ]. Comprehensibility can thus be affected. What is more, orthographic <oo> can render the sound tenser and somewhat lengthened compared to <u> (Hundleby 1963: 74–75).

Finally, the realisation of the unstressed vowel schwa /ə/ is approximated with allophones of the Xhosa phoneme /ɑ/, in particular [ɐ] or [ä]. An unstressed sound in L1 English, schwa can become stressed in a penultimate syllable on account of Xhosa phonology. The article *the* is usually pronounced [ðɛ̈] (Hundleby 1963: 80).

In summary, Hundleby identified for Xhosa-English five phonemes / i, ɛ, a, ɔ, u/ with a number of allophones, the absence of the unstressed vowel schwa as well as the absence of the distinction between long and short vowels.

Another comprehensive study was conducted by Van Rooy & Van Huyssteen (2000). They carried out the first acoustic-phonetic analysis of BSAE by using speech analysis software. In two research projects, they investigated the vowel inventory of mesolectal speakers. The first experiment analysed BSAE monophthongs and was carried out with five female Tswana-English speaking university students. The second experiment focused on the realisation of diphthongs and employed twelve students of the Sotho-Tswana group (Tswana and Southern Sotho) and the Nguni group (Xhosa and Zulu) with an equal number of male and female participants in each subsample. Their description of BSAE is based on formant frequency measurements. Table 2.1 illustrates the system of stressed vowels of Tswana-English that generally corresponds with that of Hundleby's for Xhosa-English.

Table 2.1: Major occurrence of stressed monophthongs of mesolectal Tswana-English (Van Rooy & Van Huyssteen 2000: 20)

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i] 94%	BATH, START, PALM	[a] 40%, [ɑ] 40%
KIT	[i] 81%	STRUT	[a] 59%, [ɑ] 32%
SIT[3]	[i] 50%, [ɛ] 14%	LOT	[ɔ] 70%, [ɒ] 11%
DRESS	[ɛ] 43%, [e] 29%	THOUGHT, FORCE, NORTH	[ɔ] 88%
NURSE	[ɛ] 47%, [e] 19%	GOOSE	[u] 83%
TRAP	[ɛ] 47%, [æ] 12%, [a] 10%	FOOT	[u] 100%

For TRAP, Van Rooy & Van Huysteen reported a realisation of [ɛ, æ, a], of which 47% of the tokens were produced as [ɛ], 12% as [æ] and 10% as [a]. The preference for [ɛ] distinguishes BSAE from East African Englishes that substitute the TRAP vowel with [a] (Van Rooy & Van Huyssteen 2000: 20–21). The unstressed vowel schwa is realised with the peripheral allophones [a, ɛ, e, ɑ] and central [ə]. Schwa in *the*, *a* and *an* are most frequently produced as [ɛ, e, ə] in roughly equal parts (28%, 21%, 28% of the tokens) (Van Rooy & Van Huyssteen 2000: 20–21).

There has been debate on whether BSAE is a monolithic entity or whether speakers of different indigenous African languages can be identified by the way they speak English. Van Rooy & Van Huyssteen (2000) could not observe significant differences among BSAE speakers irrespective of their ancestral languages. Hence, membership in a particular language family

[3] SIT is no lexical set in Wells' (1982a) categories, but it was used by the authors to detect a possible "KIT split", a feature that is described in Section 3.2.2.

does not play a role in the pronunciation of English. Wissing (2002) conducted a perception study on vowel quality and vowel length in different mesolectal BSAE varieties and confirmed this outcome. He suggested that at least mesolectal BSAE should not be considered a new English variety but rather an interlanguage since the speakers lack proficiency in English pronunciation. Based on Van Rooy & Van Huyssteen's findings and further analyses within the "African Speech Technology project", Van Rooy (2004) describes a mesolectal BSAE vowel system as presented in Table 2.2. Vowel contrast, characteristic of the native English varieties, is reduced by the absence of the tense/lax contrast in mesolectal BSAE as can be seen in the qualities of the vowel pairs FLEECE/KIT, NURSE/DRESS, BATH/STRUT, THOUGHT/LOT and GOOSE/FOOT. The absence of tense/lax contrast (referring to vowel quality) goes along with the absence of vowel length contrast (referring to vowel quantity). Furthermore, the contrast between central and front vowels are also reduced in that central vowels are replaced by front vowels (e.g. NURSE is pronounced as [nɛs] in mesolectal BSAE) (Van Rooy & Van Huyssteen 2000: 21; Van Rooy 2004: 945).

Table 2.2: The monophthongs of mesolectal BSAE (adapted from Van Rooy 2004: 945)

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i]	THOUGHT, FORCE, NORTH	[ɔ]
KIT	[i]	GOOSE	[u]
DRESS	[ɛ]	FOOT	[u]
NURSE	[ɛ]	commA	[ɑ̈]
TRAP	[ɛ]	lettER	[ɑ̈]
BATH, START, PALM	[ɑ̈]	happY	[ɪ]
STRUT	[ɑ̈]	horsES	[i]
LOT, CLOTH	[ɔ]	About	[ɛ ~ ə]

The absence of the contrast between phonemically long (tense) and short (lax) vowels and the frequent lack of vowel reduction in BSAE, as can be seen in the realisation of the unstressed vowels in commA, lettER, happY, horsES and About, is a transfer from the syllable-timed Bantu languages. A syllable-timed pronunciation also changes the intonation pattern of English (Gut 2005: 153–154; Van Rooy 2004: 944; Zerbian & Barnard 2008: 237).

Van Rooy (2004: 946) summarises the features of the mesolectal vowel system as follows:

- five distinct vowel phonemes /i, ɛ, a, ɔ, u/

- absence of central vowels including schwa; replacement of central vowels by peripheral vowels
- unstressed vowel in syllables other than open final ones are produced as [ɛ ~ ə]
- absence of tense/lax contrast
- preference for letter pronunciation: <o> in unstressed syllables pronounced as [ɔ]
- pronunciation of lax [ʊ] in final closed syllables between a labial obstruent and a final lateral [l] as in *double* or *careful*.

He also gives an account of acrolectal BSAE, which is closer to native South African English than the mesolect. Yet, it shows more variation within the phonemes as the mesolectal variety as Table 2.3 reveals. The basilect has not been investigated in detail yet.

The features of acrolectal BSAE can be summarised as follows (Van Rooy 2004: 948):

- eight distinct vowel phonemes, including central vowels /i, ɪ, ɛ, ɜ, a, ʌ, ɔ, u/
- two emerging but not yet established phonemes /æ, ɒ/
- unstressed vowel mostly produced as schwa
- presence of tense/lax contrast but also many exceptions.

Table 2.3: The monophthongs of acrolectal BSAE (adapted from Van Rooy 2004: 947)

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i > ɪ]	THOUGHT, FORCE, NORTH	[ɔ]
KIT	[ɪ > i]	GOOSE	[u > ʊ]
DRESS	[ɛ]	FOOT	[ʊ > u]
NURSE	[ɜ ~ ə > ɛ]	commA	[ə]
TRAP	[ɛ ~ æ]	lettER	[ə]
BATH, START, PALM	[ä ~ ʌ]	happY	[ɪ > i]
STRUT	[ʌ > ä]	horsES	[ɪ ~ ə]
LOT, CLOTH	[ɔ ~ ɒ]	About	[ə]

Reduced vowel systems are also described for other African English varieties, for example Pongweni (1990) for Zimbabwe, Schmied (1991) for East Africa and Hoffmann (2011) for Kenya. Huber (2004: 849) suggests a 5 + 2-vowel system for Ghana that comprises the phonemes /i, ɛ, a, ɔ, u/ plus /e/ and /o/ resulting from monophthongisation of the Britisch English diphthongs /eɪ/ and /əʊ/. Simo Bobda (2000a: 254) suggests the same 7-vowel system for African Englishes in general. He confirms the absence of schwa in unstressed syllables, but he argues that it does exist as an epenthetic vowel (e.g. capitalis[ə]m, macrocos[ə]m) and thus includes it into the system /i, e, ɛ, a, ɔ, o, u, ə/.

Hundleby (1963) demonstrated the difficulties in producing and comprehending L2 English sounds and suggested educational remedies. Comprehension in a technical sense was investigated by Brink & Botha (2005), who analysed the English of different Bantu-English speakers in order to determine acoustic parameters for the description of L2 Englishes. Their final goal was the adaptation of existing automatic speech recognition (ASR) systems for South African L2 English varieties. In their analysis of local TV broadcast speech, they confirmed that the central vowels /ʌ, ə, ɜ/ are rendered more peripherally as /e, ɛ, i, ɔ, ɑ/. The authors concluded that ASR systems have to be equipped with new language models in order to compensate for mergers in L2 varieties.

2.5.2 Sociophonetic studies of BSAE

This subsection introduces three studies relevant for the present work because they investigate the pronunciation of selected vowels in different South African ethnolects from a sociophonetic perspective. The researchers found features of WSAE among black speakers and conclude that ethnic labels may not always prove suitable for the denomination of language varieties.

Da Silva (2007: 115) describes the pronunciation of 76 students at the University of the Witwatersrand, Johannesburg. Her sample consisted of black and white students, divided into two genders, two mother tongue groups (L1 English and L1 Bantu languages) and three quality levels of school education based on fee range and pass rate. She investigated the lexical sets of KIT, NURSE, GOOSE, STRUT, FACE, PRICE, MOUTH and GOAT because they are "predicted to carry heavy social marking in terms of ethnicity." By means of principle components analysis (PCA), she determined two clusters of speakers, lect 1 and lect 2, according to the similarity of their linguistic behaviour. The analysis of these lects showed that lect 1 consisted of mainly white speakers who exhibited typical features of WSAE as described in the literature. Lect 2, which consisted of black speakers only, was not a homogeneous cluster but could be divided into two sublects. Typical BSAE features such as the monophthongisation of diphthongs were only produced by 51% of the speakers of lect 2. She also observed the emergence of central vowels in the speech of some BSAE speakers, especially for the KIT and NURSE vowels. Furthermore, "[t]he social composition of each lect suggests a shift from BSAE among L2 speakers of English to a variety closer to that of their white counterparts, and in some cases, contradicts the variants predicted by the BSAE variable system" (Da Silva 2007: Abstract). Da Silva therefore challenges the notion of an ethnicity-based categorisation of South African varieties of English. She argues that against the background of the changed educational conditions in particular, dialect boundaries have shifted. She therefore proposes to describe BSAE as an "L2 English lect stemming from the L1 interference of an African language" (Da Silva 2007: 240).

Mesthrie (2010b) studied the realisation of the GOOSE vowel in speakers at the top of the socio-economic spectrum. His goal was to examine whether in a post-apartheid society, young middle-class non-Whites adopt prestigious white middle-class norms – in this case GOOSE fronting, a feature that had not been reported for non-white speakers up to this point. Mesthrie's sample consisted of 48 young middle-class South Africans of the four major ethnic groups White, Black, Coloured and Indian, who had all attended fully integrated multiracial schools. He collected over 4,000 GOOSE tokens and grouped them into three categories: preceding /j/ (J-words), preceding coronals (consonants that are articulated with the flexible front part of the tongue) and preceding non-coronals. Using Watt & Fabricius (2002) ratios to determine the position of GOOSE, he reports for the black speakers:

- The formant spread, particularly of F2, is much wider compared to the white subsample.
- There is an F2 pattern according to the phonological environment.
- The non-coronals mainly range between frontish-central-backish, but have mostly central values.
- The coronals are mainly produced with central to frontish values.
- The J-words have mostly frontish or front realisation (Mesthrie 2010b: 17).

Furthermore, Mesthrie found that gender differences were relatively small. In contrast to the white group, in which males always fronted more than females (although statistically insignificant), black females produced fronter vowels than males, especially for J-words and in wordlist style. However, these differences were not significant either (Mesthrie 2010b: 18). He concludes that GOOSE fronting, an exclusive feature of WSAE so far, has become a deracialised marker of youth and middle-class status. Black females in particular exhibit the highest degree of "acculturation to the White norm", which is reflected, for example, in the fact that black and white females show no significant differences in the pronunciation of J-words (Mesthrie 2010b: 28).

The third study deals with the pronunciation of BATH which is also a social marker. Its realisation in SAE ranges between a centralised [ä:] to a fully back [ɑ:] or a rounded [ɒ:], or even backed and raised towards [ɔ:]. The centralised variant is associated with educated speakers whereas the rounded and backed forms are rather connected with speakers from the working class (e.g. Lass 2002). Mesthrie et al. (2015) compared the pronunciation of BATH among White L1 speakers of English, Coloureds, Indians and Blacks – 200 middle-class speakers in total – in the cities Johannesburg, Durban, Cape Town, Port Elizabeth and Kimberley. These cities represent the geographical spread of English in five provinces of South Africa. The survey considered social as well as regional characteristics and thus started "a pan-South African dialectology" (Mesthrie et al. 2015: 1). In each city and per ethnicity, the participants were divided into three age groups: younger speakers (18-30 years old), older speakers (over 50) and speakers between 30 and 50 years old. The acoustic analysis of around 5,500 BATH tokens elicited five variants: (1) a low and central [äː], (2) a slightly raised and central to back variant [äː] to [ɑː] or [ɒː], (3) a raised and central to back [ɑ̤ː], (4) a fully raised and superback [ɔː], (5) a raised and central [ɔ̈ː], and overlapping of these variants.

The overall result was that neither region on its own nor ethnicity on its own accounted for the variation in BATH. In terms of gender, women produced statistically significantly lower and fronter BATH vowels than men. The means of the female speakers were closer to historical prestige [äː] variant than males. However, among the Blacks, gender differences per city were by and large not obvious. Most black speakers performed a raised or slightly central to back vowel, ranging between [äː] to [ɔː]. Only in Cape Town, females exhibited a significantly fronter and lower BATH vowel than males did in their city-ethnicity-gender group. Low or back realisations among Blacks was not observed. Although region on its own was not influential, incipient regional differences among the Blacks seems to have occurred: Black speakers in Port Elizabeth and Kimberley mainly produced raised and central vowels and exhibited a lesser degree of BATH raising in the other cities.

White L1 speakers produced four variants including a fully raised superback BATH. The latter was performed by five gender-by-city groups, four of which were males. It seems that the backing of BATH is no longer a working-class feature of WSAE and obviously popular among middle-class men. The remaining speakers performed mainly an [äː] variant, in parts slightly raised. None of the white speakers exhibited a raised and central [ɔ̈ː] variant.

Mesthrie et al. (2015: 22) also observed a "bifurcation" among the Blacks. Therefore, they subdivided this ethnic group into four subgroups. One of them contained speakers "who had their entire education within segregated Black schools, and used English as L2". These speakers were labelled "traditional". A second group included speakers "who had their entire or almost entire education in desegregated schools in which White L1 norms were influential, and who speak English as a dominant language", labelled "crossover" (Mesthrie et al. 2015: 23). The difference between these two groups was statistically significant for vowel frontness, but not for vowel height. Moreover, the speech of the crossover speakers strongly resembled the features of their White counterparts per city. Crossover speakers had little in common with the pronunciation of traditional speakers.

3 White South African English

3.1 Introduction

As mentioned before, the linguistic ecology of South Africa includes beside BSAE four other major groups: White South African English (WSAE), Afrikaans English (AfkE), Indian South African English (ISAE) and Cape Flats English (CFE). AfkE is widely used in radio and television broadcast (Watermeyer 1996: 105). Therefore, it cannot be ruled out that AfkE features are found among BSAE speakers, but whether this is the case has not been tested so far. Phonetic research on AfkE is scarce, and knowledge is based on a single study of 23 Afrikaner informants from the Western Cape (Watermeyer 1996: 99). A comparison of AfkE with BSAE is tricky and is therefore not part of this work. ISAE and CFE are not considered either: ISAE is spoken predominantly in the province KwaZulu-Natal, and CFE mainly in and around Cape Town in the province Western Cape. The fieldwork for this study was carried out in the provinces Gauteng and Free State. Interaction of BSAE with those dialects are thus considered low in the region under investigation (see Figure 1.1 for a map of South Africa). Therefore, this section describes only the variety White South African English as a contact language, which is potentially the prestigious norm for BSAE speakers.

3.2 The vowel system of White South African English (WSAE)

3.2.1 Terminology and the Great Trichotomy

White South African English (WASE) refers to the L1 English variety of white South Africans of British descent. This variety has also been sometimes termed South African English (SAE) (Lass 2002; Bekker 2008) probably because it is regarded as the standard variety of the country and thus does not require any further add-ons. What is more, although SAE comprises all English native speakers irrespective of their complexion, the great majority of them are still white, and therefore both terms may be used interchangeably for the same variety. Changes in the socio-economic structure of the country, however, will lead to a steadily increasing number of non-white standard speakers of English (De Klerk 1996: 9). Hence, a clear distinction between both varieties is advisable. Since SAE in the above-mentioned sense is irrelevant for this study, this work uses WSAE to refer only to white native speakers of English (and thus neglects L1 English speakers with other ethnical backgrounds). SAE will be used as the all-embracing term for native as well as non-native English varieties.

WSAE is an "extraterritorial language" that has been brought from its "original geographical home to another area" but it principally shows the same regional variation and social stratification as the language from which it has developed (Lass 2004: 363). Based on Lanham & Macdonald's (1985) tripartite classification of sociolects that it shares with other Southern Hemisphere Englishes, Lass (2002) described WSAE as follows:

Conservative WSAE is the type of speech least distinguishable from Southern British English and based on Received Pronunciation. It is spoken by people of the upper-middle class and is the most prestigious variety. The number of Conservative speakers is decreasing (Lass 2002: 110).

Respectable WSAE derived from British migrants to Natal between 1848 and 1862. These settlers came from the Midlands and northern counties and were predominantly members of the middle and upper class (Lanham 1996: 21). This variety is regarded as the local standard covering a range of accents that are "associated with all other white standard speakers", i.e. the middle class. As the number of Conservative speakers decreases, the number of Respectable speakers is increasing and occupying "the same public sociolinguistic niche" as the former (Lass 2002: 110–111).

Extreme WSAE results from the first systematic immigration of British subjects to South Africa in 1820. About 5,000 people of largely lower-class descent from southern England settled around the Eastern Cape (Lanham 1996: 20–21). Extreme WSAE includes accents associated with speakers of the working class and those of low socio-economic status. This variety is stigmatised and is also in decline (Lass 2002: 111).

The terms 'Cultivated', 'General' and 'Broad', introduced by Mitchell & Delbridge (1965) for Australian English, are alternative denominations for the above classification and will be used for the distinction between the sociolects (see e.g. Bekker & Eley 2007).

3.2.2 Short monophthongs

KIT, DRESS, TRAP. The qualities of KIT, TRAP and DRESS are connected. They originate from the 'South African chain shift', which occurred during the nineteenth century (Lass 2002: 113, 2004: 374–376; see also Lass & Wright 1986: 208–209). The result of this shift was the centralisation of the KIT vowel [ɪ̈] due to the raising of /æ/ towards [ɛ] and original /ɛ/ towards [e] (Lass 2002: 113). Chevalier (2019: 152) observed a shift of the short front vowels, which she dubbed Reverse Vowel Shift (RVS). The RVS undoes the raising of the South African chain shift and lowers and centralises the vowels. This feature is salient among white middle-class speakers in Cape Town and other big cities of South Africa (Chevalier 2019: 168). The South African chain shift and the RVS are illustrated in Figure 3.1.

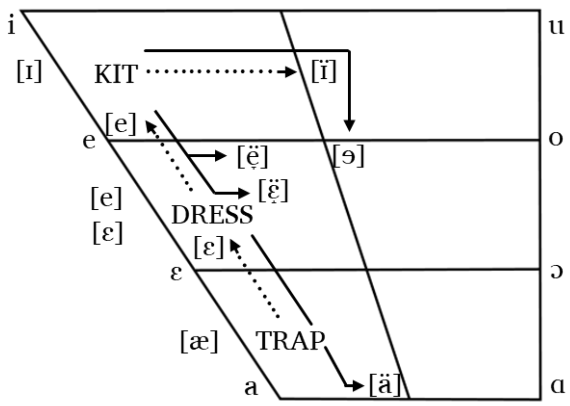

Figure 3.1: South African chain shift (dotted line) and Reverse Vowel Shift (solid line) (adapted from Chevalier 2019: 167)

Kɪᴛ. In General and Broad WSAE, the ᴋɪᴛ vowel tends to split into high front [ɪ] and a centralised [ɪ̈] variants (Wells 1982b: 612; Lass 1990: 275, 2002: 113). This distribution is generally known as the "ᴋɪᴛ split" (Wells 1982b: 612). The high front vowel [ɪ] is the marked variant and realised

- in word-initial position (*it*),
- after /h/ (*hit*),
- adjacent to velar consonants (*kids, give, sick, thing*) and
- often before post-alveolar /ʃ/ (*fish*) (Lass 2002: 115; Wells 1982b: 612).

The centralised variant [ɪ̈] is the unmarked variant and occurs in any other environment. It is often further retracted to [ə] in the vicinity of labials, such as *lip, miss, tin* and *slim* and after /r l/ (Wells 1982b: 612; Lass 1990: 275, 2004: 374). In Broad WSAE, the fronter ᴋɪᴛ allophone is raised and fronted further to ꜰʟᴇᴇᴄᴇ (Lass 2002: 115; Wells 1982b: 612). Here, *kit* and *bit* do not rhyme, for which reason Wells (1982b: 612) suggests the variants [ɪ~i] as allophones of /ɪ/ and the variants [ɪ̈~ə] as allophones of /ə/.

Bekker (2008: 268), by contrast, does not regard it a phonemic split but rather an allophonic variation of the phoneme /ɪ/. He investigated the pronunciation of young middle-class female WSAE speakers and found a "virtual wholesale centralization of ᴋɪᴛ" with [ɪ̈]-like pronunciation for fronter allophones and [ɘ]-like values for more central allophones (Bekker 2008: 277).

The KIT split is a salient social marker. In Cultivated WSAE, it is practically absent, that is, *it* rhymes with *sit*, but the value of [ɪ] is relatively front. This form is also used by General speakers in more formal styles and particularly by women (Lass 2004: 375). Generally, however, General speakers use [ɪ] in *it* and [ɪ̈] in *sit*, and Broad speakers use [i] for *it* and [ɪ̈] for *sit* (Lass 2002: 114; Bekker 2008: 251).

In syllable-final /l/ (dark or velarised [ɫ]) and in post /w/ contexts, KIT can merge with FOOT resulting in homophone pairs, such as *woman/women* or *bill/bull* with a KIT value of [ɤ̈] (Lass 1990: 275, 2002: 115). Bekker (2008: 269–270) also found distinct retraction of KIT before tautosyllabic dark [ɫ] in *till*, *fill* and *pill*.

DRESS. The DRESS vowel /ɛ/ in WSAE is realised as [e] and is produced closer by women than men. Before dark [ɫ], DRESS is retracted to [ɛ] or even [æ] in Broad WSAE and some General varieties (Lass 2002: 115). Bekker (2008: 428) found a "nearly-invariant" overlap of DRESS and the fronted KIT allophone.

TRAP /æ/ is an important social marker since there is a clear distinction between [æ] among Cultivated and General speakers and [ɛ] among Broad speakers. Generally, the TRAP vowel lowers and retracts before dark [ɫ] (Lass 2002: 115). Bekker & Eley (2007) investigated the pronunciation of monophthongs in citation form of ten white young Cultivated females. They found a lowering and retraction of TRAP and assume it a prestige form as a "hypercorrect movement away from the [ɛ] of Broad WSAfE" (Bekker & Eley 2007: 113). In his analysis of young middle-class females, Bekker (2008: 428) found a lowering of TRAP and assumes an endogenous change in WSAE.

Table 3.1: The short vowels of WSAE according to Bowerman (2004: 936)

Lexical set	Allophones
KIT	[ɪ]~[ɪ̈], [ə], [i]
DRESS	[e]
TRAP	[æ] > [ɛ]
LOT	[ɒ̈] > [ɔ], [ʌ]
STRUT	[ä] > [ɐ]
FOOT	[ʊ] > [ʉ]

LOT, STRUT, FOOT. LOT is often realised as a centralised allophone [ɒ̈] or with a quality similar to [ʌ]. STRUT is produced around [ɐ̶] and [ä]. Backer and more open variants are associated with Cultivated and General speakers. FOOT is realised as a fairly-close rounded vowel [ʊ] or slightly fronter in all varieties (Lass 2002: 115). Table 3.1 gives an overview of WSAE short vowels.

3.2.3 Long monophthongs

FLEECE is produced as a long close vowel in all varieties (Lass 2004: 376).

NURSE is a mid central unrounded vowel [ɜː] in Cultivated. In General and Broad, it is a close-mid front rounded [ø], [øː] or slightly lower. Here, the quality is similar to the vowel in German *schön* or French *peu* (Lass 2002: 116, 2004: 376).

BATH /ɑː/ is a social marker. Cultivated speakers realise a centralised [ɑ̈ː] or a central [a̠ː]. General and Broad speakers produce an unrounded back vowel [ɑː]; Broad speakers may produce a rounded allophone [ɒː] (Lass 2002: 116, 2004: 377; Bekker & Van Rooy 2015: 292) or even an [ɔː] variant (Mesthrie et al. 2015: 1).

THOUGHT vowels are usually produced as either [ɔː] (Cultivated) or [oː] (General and Broad). If THOUGHT is followed by a voiceless fricative, it may be pronounced as LOT (Lass 2002: 116, 2004: 377).

GOOSE is an important social marker. It is produced by Cultivated speakers as a rounded "back(-ish)"[4] vowel [uː]. General and Broad speakers produce a centralised allophone [ʉː] or fronter. Younger General and Broad speakers, in particular females, may produce a rounded FLEECE vowel [yː] (Lass 2002: 116, 2004: 377; Mesthrie 2010b: 12). Table 3.2 provides as overview of WSAE long vowels suggested by Bowermann (2004). There is apparently no phonological distinction between BATH, PALM and START since the latter two are not considered in the list. The same applies to THOUGHT, NORTH and FORCE of which only THOUGHT is listed.

Table 3.2: The long vowels of WSAE according to Bowerman (2004: 936)

Lexical set	Allophones
FLEECE	[iː]
NURSE	[ɜː]
GOOSE	[uː], [ʉː] > [yː]
THOUGHT	[ɒ̈] > [ɔ], [ʌ]
BATH	[ä] > [ɐ]

[4] Mesthrie (2010b) used the term *backish* to describe a vowel that is produced halfway the dimension central vowel and back vowel, adapted from Watt & Fabricius (2002).

4 Dataset and Methodology

4.1 Recruitment of informants

Data collection was carried out in the Republic of South Africa within a six-week period of fieldwork between 7th August and 9th September 2013. The participants in the survey were exclusively black South African citizens whose mother tongue was a language indigenous to South Africa. The main site of data collection was the Vaal Triangle Campus of North-West University (NWU) in Vanderbijlpark in the Gauteng province, about 70 km south of Johannesburg. With 9.49 million black South Africans, Gauteng has the biggest proportion of this ethnic group per province in South Africa (Census 2011: 21). This proportion is also reflected in the composition of the student numbers at NWU: Of the 6,512 students matriculated in 2013 (North-West University 2015: 25), the majority was black (Coetzee-Van Rooy 2012: 92). This demographic situation was a promising basis for successful data collection. A second interview venue was a kindergarten in Phiritona, a township in the Free State province, 73 km south of Vanderbijlpark.

The initial intention of this apparent time study was to compile a stratified sample (Schilling-Estes 2007: 169) comprising data of black South Africans from three age groups (young, middle and older). The age difference between these groups should have been sufficiently large to obtain data from three generations. Therefore, the prospective participants were supposed to be around 20 (18–24 years old), 40 (38–43 years old) and equal to or older than 60 years old. By choosing these age groups, it would have been able to detect possible language differences between three generations and thus a possible language change in progress (Labov 1994: 45–46).

The young participants were exclusively students. They were contacted randomly on campus, and appointments were made with those interested in participation. Staff members of NWU also propagated the study in their classes and among their colleagues. Despite the students' genuine willingness to provide their names and phone numbers, and despite the honest excitement of many of them in participating in the study, it was challenging to find people who actually came to the interview session. Verbal consent to participate or even a fixed appointment was no guarantee for appearance. Therefore, schedules and the planning of fixed interview slots were frequently subjected to improvisation.

Despite this difficulty, recruiting people for the age group 'young' was comparatively easy because of the large number of students on campus. It was more problematic to find participants for the age group 'middle' and 'older'. Although it was possible to find many people who were willing to participate in the study, finding a sufficiently large number of participants that fitted within the age group boundaries was not successful. The biggest problem was to find participants for the category 'older' of which four were interviewed on campus. Expanding the search for participants geographically was given up due to time constraints, lack of opportunity (with one exception) and, admittedly, safety concerns. In consequence, the sampling strategy had to be changed from the collection of a stratified sample to that of a convenience sample (Bryman 2012: 201), that is, any person who was willing to participate was chosen. A convenience sample, however, may cause problems in terms of representativeness and generalisation of the data (Rasinger 2013: 51).

With regard to the way to find informants, a "friend of a friend" (or "snowball") technique as successfully utilised by many researchers (e.g. Milroy & Gordon 2003: 32–33; Schilling-Estes 2007: 179) worked only sporadically. Schilling-Estes describes this method as "far less intimidating and unnatural than simply walking up to people one doesn't know" (2007: 181). However, researchers have also successfully gathered data also by "simply hanging out in neighbourhood streets" (Schilling-Estes 2007: 180). The hanging-out-on-campus method to contact students proved successful. Da Silva (2007) took a similar approach with her comparison of language behaviour among black and white students.

4.2 The sample

The original sample consisted of 50 speakers. Some of them had limited vocabulary so that they gave short answers, used the same phrases repeatedly and were unable to dwell on a topic. These participants had stated in the questionnaire that they left school prematurely. In order to obtain sufficiently diverse material from each participant, only speakers who had completed secondary school were selected. Of the 50 participants recorded, this applied to 44 speakers. Table 4.1 and Table 4.2 show the composition of the sample according to home language, language family, gender and age group.

The distribution of the languages reflected the general language distribution of the Vaal Triangle region where most recordings were carried out. The Vaal Triangle region belong to the Gauteng province, and 36 participants stated that they currently live in Gauteng. As reported in the 2011 national census (25–26), the major home languages for this province (including the Johannesburg conurbation) are Zulu (19.8%), English (13.3%), Afrikaans (12.4%), Southern Sotho (11.6%) and Northern Sotho (10.6%). These languages are spoken by roughly 67.7% of the population. In the Vaal Triangle region, however, Afrikaans and Southern Sotho are the dominant languages (Coetzee-Van Rooy 2012: 92). Since a prerequisite of BSAE is an L1 Bantu language, Afrikaans was no home language of the participants investigated. The majority of the informants (29) spoke Southern Sotho as their mother tongue. The language families formed the basis for the division of the sample into L1 with 5 vowels (Venda and Nguni languages) and L1 with 7 vowels (Sotho-Tswana languages) (see Section 4.4.5).

Table 4.1: Composition of the sample by first language and language family

L1 / Language family	Nguni	Sotho-Tswana	Venda	Total
Ndebele	1			
Xhosa	8			
Zulu	5			
Northern Sotho		1		
Southern Sotho		21		
Tswana		7		
Venda			1	
Total	**14**	**29**	**1**	**44**

Table 4.2: Composition of the sample by gender and age group

Gender/Age group	Young (19–23)	Middle (25–54)	Older (58–84)	Total
Male	9	10	2	21
Female	11	7	5	23
Total	**20**	**17**	**7**	**44**

4.3 Social variables

Linguistic variation correlates with social variables, such as age, gender, socio-economic class and ethnicity (Labov 2010: 197). The following explains the inclusion of specific variables and the exclusion of others.

4.3.1 Age

Second language proficiency strongly depends on the age of the learner. Very young children are well-adapted for language learning while older children or adults have more difficulties to acquire language, a phenomenon that is related to Lenneberg's (1967) critical period hypothesis (Lenneberg 1967; see also Johnsons and Newport cited in Harley 2014: 76).

Second language proficiency is also determined by the quality and quantity of the input, that is, the proficiency of the teachers, the quality of the teaching material and the frequency of contact to native speakers (see also Section 2.1). The fieldwork for this study was conducted in 2013. At that time, the youngest participants (between 19 and 23 years old) had been able to experience an officially unsegregated society from a very early age on. This circumstance could have included very early exposure to L1 English. Against the background of the drastic political, social and economic changes in the country, age is assumed to play a major role in the outcome of this study. The sample was thus divided into three age groups: young (exclusively university students), middle and older.

4.3.2 Gender

The variable gender was included because linguistic behaviour often follows a certain gender pattern. Language change, for example, is reported to often be led by women. Generally, women use fewer stigmatised forms, and they favour prestige forms (change from above[5]) more often than men (Labov 1991: 243, 1994: 156, 2001: 366). Lanham & Macdonald (1985: 29) confirm: "Women of all ages are more assiduous in their pursuit of what they perceive as prestige norms, and are likely to be more advanced in phonetic speech." Dubois & Horvath (2000: 288) suggest that "women are more sensitive than men to the social evaluation of speech and use more of the positively evaluated variants and less of the negatively variants than do men." Assuming that a white English variety is the aspired norm for many Blacks and that females adopt prestigious forms faster than men, it is expected that the cohort investigated shows the pattern suggested by Labov, that is, women, in particular young ones, may show a higher degree of adaptation of a prestigious form than men.

[5] Change from above is the result of social factors that influence language use. It usually involves the percolation of high-prestige forms from the highest social class to lower ones, but it also includes the diffusion of elements from other language systems, e.g. from other non-standard dialects (Labov (2010: 185).

4.3.3 Ethnicity and first language

Ethnic membership of the informants was not inquired because it was more important to elicit the language(s) they acquired first and predominantly use, which can differ from the membership of an ethnic group by blood. Apart from that, in inter-ethnic marriages, the children belong to two ethnic groups but may be raised in the tradition and language of just one of them. Furthermore, people of one ethnic group may adopt a language that is not their home language due to the circumstances of the immediate environment. Four examples shall illustrate the manifold ethnic and linguistic constellations in the country:

1. Speaker TMM12's father was a Shona from Botswana and came as a migrant labourer to South Africa. Due to the apartheid policy at that time, he had to change his name into a name registered in South Africa. He chose a Tswana name and was thus assigned to Bophuthatswana, the homeland of the Tswana ethnic group in the north-western region of the country. In this Tswana-speaking area, he adopted Tswana as his preferred language and therefore, the mother tongue of speaker TMM12 became Tswana.
2. Speaker VOM01's mother was Venda and his father Northern Sotho. His first language was Venda. He stated that he could speak Zulu, Tsonga, Northern Sotho and other languages equally fluently. He saw the reason for his multilingualism in the fact that he grew up in a mining town bringing together people from different language backgrounds.
3. Speaker TYF01 explained her language situation as following:

> Like my father, he's Xhosa. My mom is Sotho, so that means I'm Xhosa as well. But at home we don't speak Xhosa. We speak Setswana because I grew up in an environment where they speak Setswana. We speak Setswana at home, we don't speak Xhosa at all.

The informant reported to speak also English and Afrikaans but none of the home languages of her parents.

4. Speaker OYF13 stated Xhosa as the first language she acquired, but explained:

> We are Tsongas, I'm a Tsonga, I'm a pure Tsonga. […] and then I studied […] Southern Sotho, and then […] in high and primary school […] did IsiXhosa. So, right now I'm perfectly speaking IsiXhosa.

This speaker stated that she also learnt Zulu, Southern Sotho, English and Tsonga. The question "Which language do you speak best?" (see Section 5.1), she answered with Xhosa and Southern Sotho, indicating that she has no native-speaker competence of Tsonga, the home language of her parents. Although it can be assumed that her first utterances were in Tsonga, the informant chose not Tsonga but Xhosa as the language she first acquired.

4.3.4 Class and education

The variable (socio-economic) class, is a complex composition of status features, including education, occupation, income, neighbourhood and housing (Trudgill 1974). Numerous researchers have found a correlation between language behaviour and class membership, but the definition of "class" can pose problems even in socially stable societies (Milroy & Gordon 2003: 98). In the context of South Africa, *class* is even more difficult to define. One reason is the high degree of social mobility among the black population. Another problem is the categorisation of students. Most students investigated came from low-income families and could thus be regarded as members of the working class. However, a university student is also always a prospective white-collar worker. Therefore, they could equally be categorised as members of the middle class. Coetzee-Van Rooy (2012: 92) investigated first-year university students in the very same area where the present data collection took place and categorised her sample as largely "upwardly mobile urban working class". This description is also applicable to the young participants of the present sample.

Data was collected from an availability sample, for which reason it was impossible to control for class membership, and *education* was instead the decisive variable. Interviews were conducted with participants of different educational backgrounds, but the baseline for the analysis was their completion of secondary school. An overview of the speakers' metadata is provided in Appendix A-1.

4.4 Linguistic variables

4.4.1 Phonemes

The phonemes under scrutiny are stressed monophthongs. They are labelled with the system of the lexical sets proposed by Wells (1982a). The lexical sets are FLEECE, KIT, DRESS, NURSE, TRAP, BATH, START, PALM, STRUT, LOT, CLOTH, THOUGHT, FORCE, NORTH, FOOT and GOOSE. For allophonic representations within the lexical sets, IPA symbols are used.

4.4.2 Phonological context

The phonological context can influence formant values and duration (e.g. Thomas 2011: 49; Nycz & Hall-Lew 2014: 4). Especially the features of neighbouring consonants like manner or voicing can have a very strong effect on vowel behaviour. Therefore, the environment adjacent to the vowel was included. A detailed explanation of the inclusion and omission of certain contexts is given in Section 4.7.1.

4.4.3 Spelling

Spelling variants of vowels may have an influence on both vowel quality and quantity. This has been attested for the quality of NURSE in various African Englishes (Simo Bobda 2000b) and for the quantity of FOOT in Xhosa-English (Hundleby 1963). The lexical set STRUT has also a number of graphemic representations. Their influence on vowel quality has not been tested yet. The variable *spelling* was included to investigate the influence of spelling variants of NURSE and STRUT on vowel quality and those of FOOT on vowel duration.

4.4.4 Speech style

Acoustic data was collected by conducting sociolinguistic interviews (Labov 1984b). A traditional sociolinguistic interview consists of three parts: a free-flowing conversation/interview (IN), reading a short reading passage (ReP) and reading individual words from a wordlist (WL). These three parts aim to obtain data with different degrees of formality: talking/mostly informal (IN), reading style/more formal (ReP) and citation style/most formal (WL). Sections 4.6.2 and 4.6.3 describe the setup of these interview parts.

4.4.5 Vowels in L1

The transfer of linguistic features from the mother tongue is particularly noticeable in the accent. BSAE sub-varieties resulting from the influence of different substrate languages have not been reported so far (Van Rooy & Van Huyssteen 2000; Wissing 2002; Louw & Wet 2007), and Van Rooy & Van Huyssteen (2000: 30) describe BSAE as "a fairly coherent variety of English". As some participants claimed to be able to distinguish the English accents of speakers of the Bantu subgroups Nguni and Sotho-Tswana, it was decided to test the influence of the mother tongue on BSAE. However, the uneven distribution of the participants' home languages would have made the interpretation of the results difficult. Therefore, they were divided according to the vowel systems of the substrate languages. Thus, the variable *number of vowels in L1* was tested with the variant '5 vowels' (Venda and Nguni languages) – 15 participants – and the variant '7 vowels' (Sotho-Tswana languages) – 29 participants.

4.5 Research hypotheses

The aim of this apparent time study is the documentation of a possible language change. The consideration of the social and linguistic aspects described above should contribute to this aim. The following research hypotheses are thus put forward.

With a new policy after 1994, it is assumed that access to native English (in whatsoever way) has increased – a situation the younger generation has benefited most because they grew up after Event X. Therefore, age is considered a decisive factor:

H1: The phonologies of the age groups are significantly different.

H1.1: The youngest age group differs significantly from the other two.

H1.2: The middle and older age group differ from each other but to a lesser extent than to the youngest age group.

H1.3: The younger the age group, the higher the number of distinct vowels.

While the influence of age is linked to the political changes in the country, the variable gender is connected with the role and status of men and women in society. The language of the sexes is reported to be different in word choice and pronunciation in that women tend to prefer prestige norms. Since this seems to be a universal property of speech, the following hypotheses are thus formulated:

H2: Gender differences within the age groups are significantly different.

H2.1: Young females differ significantly from the other participants.

The next hypothesis refers to the vowel systems of the substrate language families. Although so far, they have been reported to have no influence on pronunciation, they were considered nevertheless:

H3: Speakers whose native language has a seven-vowel system have more distinct vowels in English than those who have a five-vowel system.

Finally, intra-speaker variation due to speech style can occur in that speakers adjust their pronunciation according to the level of formality. Hence:

H4: The vowels in the speech styles show significant differences.

The statistical model used for hypothesis testing is described in Section 4.7.3. The significance level assigned is $\alpha = 0.05$. The hypotheses will be accepted if the probability value $p \leq \alpha$.

4.6 Research design and recordings

The participants were recorded while reading words from a wordlist and a short written passage and during a conversation. Some scholars suggest an order that starts with the most informal speech style (conversation) and ends with the most formal one (wordlist) (Krug & Sell 2013: 73; Trudgill 1974: 46). Other studies suggest that the order of the three parts is irrelevant because the people are able to choose from a repertoire of styles and adapt their style according to the needs of the communicational situation (Milroy & Gordon 2003: 50–51; Schilling 2013: 332). This may become normal practice as people increasingly come in contact with more language varieties (Schilling 2013: 332), which fully applies to the linguistic setting of South Africa. Since there is no serious objection to any order, the recording started with the interview, followed by the reading passage and ended with the wordlist. It seemed more convenient for the interviewee to be involved in a casual conversation first to accommodate him or her to the situation.

Of the forty-four participants chosen for further investigation, forty took part in all three parts of the sociolinguistic interview. Four participants took part only in the interview (IN).

Thirty-two participants were recorded in individual interviews. Twelve participants were recorded in pairs because they wanted to be interviewed together with a friend. In order to create a pleasant atmosphere for them, the researcher consented to this request. Thirty speakers were interviewed on the premises of the Vaal Triangle Campus of NWU in an office that had no outer walls, no windows and a sound-attenuating door. Although this room was not completely sound-proof, it proved to be a big advantage in terms of closing off outside noise. Other venues were less favourable, yet the recordings are of high quality. At the beginning of each recording session, the participants were asked to sign an informed consent form and to fill in a questionnaire providing demographic data and information on their language skills as well as on their usage of and attitudes towards English (Appendix A-2).

4.6.1 Coding of the speakers

The speakers were coded by ethnicity, gender, age group and a running number. The codes for ethnicity were Ndebele (D), Northern Sotho (N), Tsonga (O), Southern Sotho (S), Tswana (T), Venda (V), Xhosa (X) and Zulu (Z). Gender was coded male (M) and female (F), and age group with young (Y), middle (M) and older (O). For example, SMM10 signifies a middle-age male Southern Sotho speaker with the running number ten; TYF04 signifies a young female Tswana speaker with the running number four.

4.6.2 The interview

The interviews lasted between 25 minutes and 65 minutes. The single interviews were 25–48 minutes long and the pair interviews 40–65 minutes. The questions based on conversational modules and included the topics about work/school, family and friends, childhood memories, games, religion, crime etc. (Labov 1984a: 33–35; Tagliamonte 2006: 37–49; Milroy & Gordon

2003: 57–68). The interviews, however, could be considered as semi-structured because the interviewer only partly followed a predetermined catalogue of questions. For example, when the interviewee was a staff member, information from the person's website was gathered beforehand in order to prepare precise questions on his or her work and current projects.

The interviews aimed to obtain the least formal speech style, ideally the one used with friends. However, the linguistic behaviour of the informants may differ from their normal use because of the presence of an observer, a phenomenon that is known as the Observer's Paradox (Labov 1991: 209). Despite all efforts to make the sessions for the informants as comfortable as possible, an interview remains a very special situation in which the informants may not react naturally. Therefore, the interviewer can never be sure whether the vernacular is used or not. This may apply in particular when the ethnic membership of interviewer and interviewees is apparently different as this was the case in the present study. However, Trudgill (1974: 46) suggests that even if respondents feel more constrained and thus speak more formal than other respondents, they will nevertheless exhibit different levels of formality during the interview, in the reading passage and in the citation style. What is more, the mother tongue of the informants were African languages. It is therefore most likely that their vernacular – if there is only one vernacular (Schilling 2013: 332) – is a variety of their mother tongue rather than of English. Therefore, the three sections of the sociolinguistic interview may elicit different speech styles, but most probably not the vernacular of the speakers. In this context, it is worth mentioning that the description of and recommendations for the setup of sociolinguistic interviews mainly refer to monoethnic and monolingual environments. Applying such a research method to a multilingual African setting may thus be a methodological issue.

4.6.3 Reading Passage and Wordlist

The reading tasks included the reading of a text passage of 221 words and 125 individual words. The informants read the instruction for these tasks on a laptop and could decide themselves when to start the task by tapping on the trackpad of the laptop. After a 3-2-1 countdown, a slide with the story "The boy who cried 'Wolf'" (Deterding 2006) appeared, which the participants were asked to read aloud. This text was chosen because it has some advantages compared to other texts for phonetic research: It is sufficiently long to contain at least three tokens of each English monophthong (except of the lexical set PALM). These monophthongs can be easily measured because the boundaries to the adjacent environments are easy to detect. The text also offers a great variety of phonological contexts as well as five minimal pairs. Although not every context was considered for the present analysis, the remaining tokens build a solid basis for further exploration.

At the end of the reading passage, the participants were given further instructions for reading out individual words from a wordlist. Again, tapping the trackpad started a second 3-2-1 count-down for the wordlist. Each slide showed one word and changed automatically after two seconds. The wordlist was designed to include a wide range of consonantal environments for each vowel. The participants needed approximately six minutes altogether to accomplish both reading tasks. The reading passage (ReP) and the wordlist (WL) are shown in Appendix A-3 and A-4.

4.6.4 Recording device and data storage

The recordings were carried out with Zoom H4n Handy Recorder and one or two lavalier microphones (Audio-Technica AT831b) depending on the number of participants in the interview. The advantage of a lavalier microphone over a table microphone or a headset is that it is small and can be flexibly attached to the clothes. The distance between microphone and the speaker's mouth – preferably 15 cm – stays practically the same, even when the participant moves (Labov 1984a: 33). Although an interview is anything but a natural setting, a small and unobtrusive device may at least partly reduce the participant's consciousness of being recorded.

Most recordings were made at a sampling frequency of 48,000 Hz and a 32-bit quantisation. This setting applied to all occasions, in which the Zoom recorder could be connected to a power supply. On occasion when the recorder had to be used with batteries, the device only accepted a sampling frequency of 44,100 Hz and a 16-bit quantisation. The lower sampling rate was still high enough to yield high quality recordings. The recordings were stored in the audio file format *.wav.

4.6.5 Data preparation for the acoustic analysis

The acoustic analysis of the sound samples was carried out in Praat, version 5.3.63 (Boersma & Weenink 2014). An acoustic analysis with Praat requires an audio file as the acoustic representation of speech and a corresponding TextGrid for the annotation of the speech segments. TextGrids can be directly generated in Praat, but in the present study, they were produced in a more time-saving manner by means of different software tools.

In a first step, the audio files were orthographically transcribed. In the case of ReP and WL, this transcription was carried out in the format *.txt and was identical for all participants who read both parts. In the case of IN, the orthographic transcription was carried out with ELAN (**E**UDICO **L**inguistic **AN**notator, version 4.9.1), a tool for the annotation of video and audio files (ELAN 2015). ELAN offers several features that increase the speed and efficiency of the laborious transcription work. For example, it is possible to control the playback rate allowing the researcher to activate a lower speed if necessary. The segmentation is time-aligned, which is useful when searching for a specific frame. It is also possible to create a separate audio tier for each speaker and thus unambiguously separate two speakers in a pair interview.

ELAN was used exclusively for the orthographic transcription. Therefore, the segments did not consist of single speech sounds but longer stretches of speech – from words to whole sentences. The segmentation did not follow a certain scheme, but segment boundaries were usually set during a visible speech pause as can be seen in Figure 4.1. Since only the speech of the interviewees was relevant for the study, the utterances of the interviewer were not transcribed. Therefore, interviews with one participant consisted of one tier and interviews with two participants of two tiers.

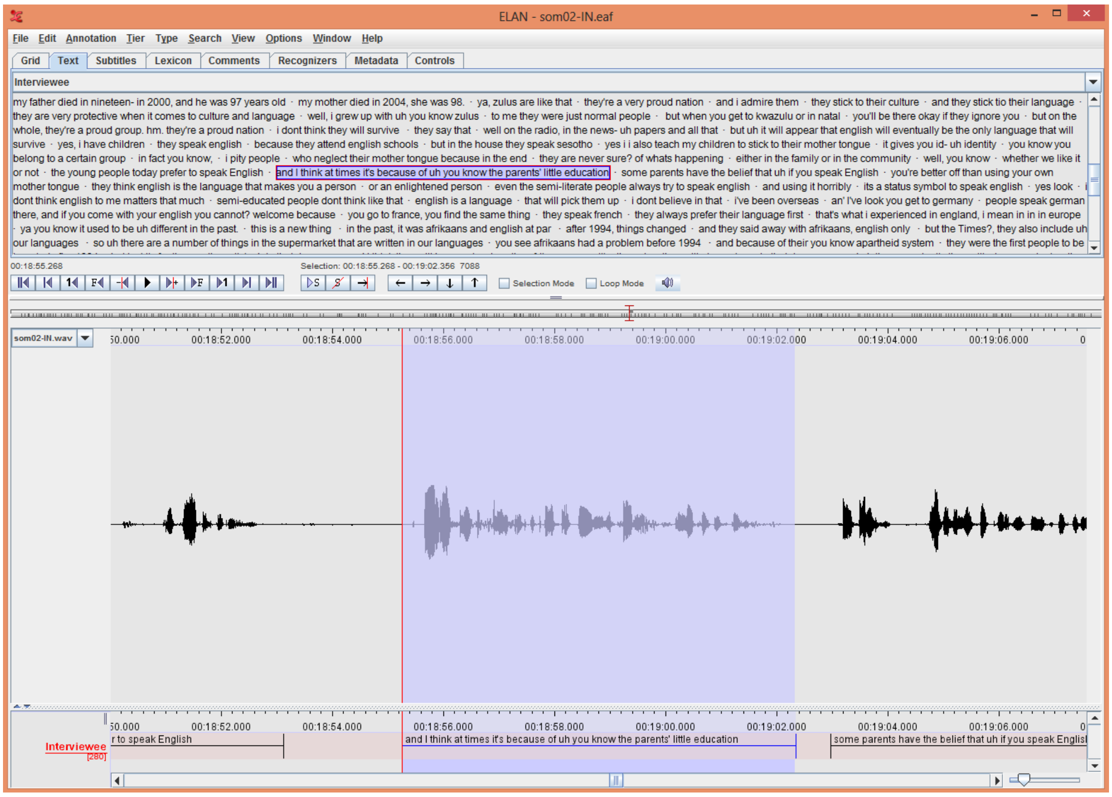

Figure 4.1: Orthographic segmentation for an interview with one speaker in ELAN

The ELAN files are created as *.eaf (EUDICO Annotation Format), a format that cannot be processed by Praat, but the programme can export the transcribed files as Praat TextGrids. At that stage, a Praat TextGrid for one speaker consisted of text segments (the transcribed utterances of that very speaker) and blank intervals (those parts in which the interviewer or a second speaker talked, or parts that had not been transcribed, such as laughter, coughs, interjections, overlapping speech etc.).

The respective audio file contained segments with the utterances of one speaker, but also segments with utterances of the interviewer or a second speaker. The utterances of the latter two as well as untranscribed passages (such as laughter etc.) correspond with blank TextGrid segments. For further processing, the audio files had to be cleared of all irrelevant utterances and blank spaces. Therefore, with the help of several Praat scripts, every section in the audio file was removed where the respective TextGrid segment was blank. These scripts also replaced every removed section with an interval of 1 second and adjusted the blank sections in the TextGrid accordingly. The preliminary outcome of the Praat-processed ELAN TextGrids are txt files containing orthographically transcribed utterances exclusively of a single speaker in chronological order. During the transcription and segmentation process in ELAN, efficiency was given priority over grammatical correctness. Therefore, typos were not corrected immediately in ELAN (which has no spell checker) but by means of the Praat-generated txt files.

The audio files (in *.wav format) and the txt files were used for the final segmentation. This step was carried out with MAUS (**M**unich **AU**tomatic **S**egmentation), a software package for the automatic segmentation and phonetic transcription of audio files (Schiel et al. n.d.; Schiel 1999). The programme aligns sound signals to the transcript of these signals, produces single sound segments and allocates them with the phonetic symbols of SAMPA (**S**peech **A**ssessment **M**ethods **P**honetic **A**lphabet) (Wells 1997).

MAUS can segment speech in several languages. For English, it provides four English L1 varieties, labelled 'Australian', 'Great Britain', 'New Zealand' and 'American'. To verify which annotation mode would be most suitable for the present analysis, a test run was carried out producing four TextGrids in four different annotation modes of the same text. A comparison showed that none of them were entirely appropriate for further use, and all of them needed to be modified in equal measure. Therefore, the mode 'Great Britain' was chosen because of the colonial background of South Africa.

The SAMPA symbols were transformed into a suitable format. This included the rearrangement and labelling of the tiers as well as the labelling of the sound segments where necessary. For the labelling of vowels, the lexical sets proposed by Wells (1982a) were used. Their advantage to the symbols of the IPA or the convention used by Trager & Bloch (1941) or Labov (1966) is that the lexical sets embrace all English vowels irrespective of the variety in which they occur. For instance, GOOSE vowels may be pronounced differently in different English accents (short, fronted, lowered, etc.), but they still belong to the very lexical set GOOSE. The use of lexical sets has also a very practical benefit: They can be written much faster than IPA symbols.

As mentioned in Section 2.5.1, due to the transfer of prosodic features from the substrate languages, African Englishes, especially L2 varieties, are often syllable-timed, meaning that vowels, which are unstressed in L1 varieties may be unreduced in vowel quality and quantity. This was also the findings of Van Rooy & Van Huyssteen (2000). In their analysis of BSAE vowels, they distinguished between monophthongs and the unstressed vowel schwa /ə/. For the latter, they described at least seven allophones, most of which were produced as full vowels. For the present study, only monophthongs which are typically stressed in native English varieties were analysed.

4.6.6 Segmentation

The segmentation and acoustic analysis of the vowel tokens was carried out in Praat. The standard formant and spectrogram settings were generally adopted, that is, the maximum formant value was 5,000 Hz for male speakers and 5,500 Hz for female speakers. The dynamic range of the spectrogram was changed to 50.0 dB because the default setting of 70.0 dB often produced a very dark spectrogram with little contrast. The formant and spectrogram settings are seen in Table 4.3.

Table 4.3: Formant and spectrogram setting in Praat

Formant parameter	**Setting**	**Spectrogram parameter**	**Setting**
Maximum formant value	5,000 Hz (male) 5,500 Hz (female)	**View range**	0.0 – 5,000 Hz
Formants	5	**Window length**	0.005 s
Window length	0.025 s	**Dynamic range**	50.0 dB
Dynamic range	30.0 dB		

The adjustment of the segment boundaries by MAUS was very timesaving. Yet, in almost all instances, the boundaries had to be readjusted because the demarcation by MAUS was incorrect to various degrees. An extreme example is shown in Figure 4.2.

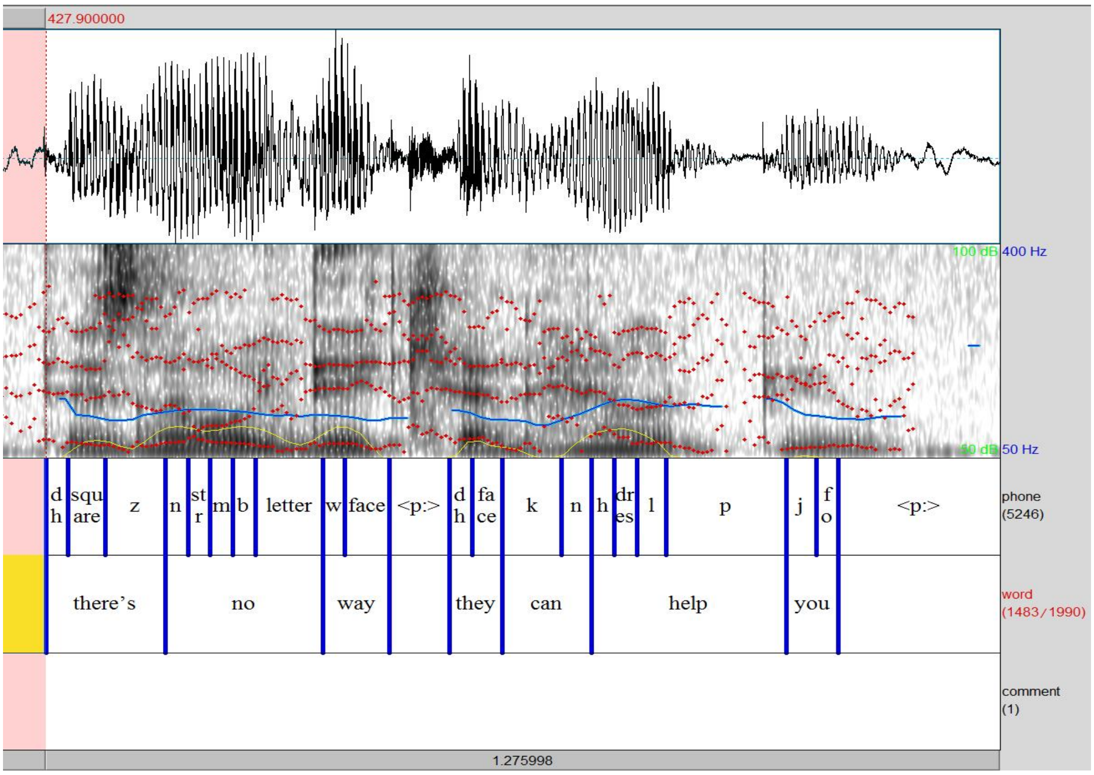

Figure 4.2: Praat window with waveform (upper panel), spectrogram (second panel) and three annotation tiers

It contains a number of mistakes, three of which shall be explained: First, what is categorised as a soundless pause <p:> (in the centre of the figure), is actually phoneme /k/ of the word *can*. Second, the word *no* was identified as the abbreviation of *number* and was split into five segments |n|strut|m|b|letter| instead of two |n|goat| and is displayed longer than it was actually uttered. Third, the word *you* contains the segment FOOT ('fo'), indicating a short vowel, instead of the long vowel GOOSE. Because of the partly flawed MAUS segmentation, all relevant segments were manually checked for accuracy using the principles suggested by Machač & Skarnitzl (2009) illustrated in Figure 4.3.

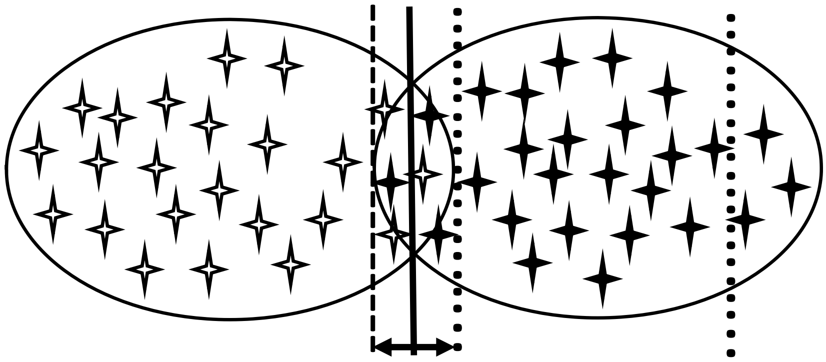

Figure 4.3: Options for boundary location between neighbouring segments (adapted from Machač & Skarnitzl 2009: 18)

There are three options for the location of the boundary between two segments relevant for this study:

1. If two segments are of equal importance, the boundary is placed in the middle of the transition area (Figure 4.3, solid line).
2. If one segment is dominant (e.g. the filled asterisks), the boundary is placed outside of this segment (Figure 4.3, dashed line).
3. If the centre of one segment is dominant (e.g. the filled asterisks), the boundary is placed outside the centre of this segment (Figure 4.3, dotted line) while peripheral elements remain beyond the boundary (Machač & Skarnitzl 2009: 19).

For this study, only option 2 was relevant. Machač & Skarnitzl (2009) also propose three general segmentation rules:

1. Boundaries are placed between two glottal pulses.
2. In a transition phase (an area in which a boundary cannot be unambiguously placed) it is placed in the temporal midpoint of this area.
3. Vowel boundaries are always placed at a zero crossing (Machač & Skarnitzl 2009: 23–24).

These rules were fully adopted. However, particularly in transitional areas, the demarcation was frequently decided by a multi-dimensional consideration of waveform, spectrogram and auditory impression of the vowel sound. Manual placement of boundaries “sometimes involves a great deal of human judgment” (Peterson & Lehiste 1960: 694) and “intellectual activity” (Machač & Skarnitzl 2009: 18), and a strict adherence to the above-mentioned suggestions should guarantee consistent segmentation.[6]

[6] For a detailed description of segmentation cues, see for example Machač & Skarnitzl (2009) and Peterson & Lehiste (1960).

4.6.7 Data cleansing

The quality of a vowel sound is determined by its acoustic resonances in the vocal tract. These resonances are called formants and are the points of maximum energy in the spectrum of a speech sound. Most vowels can be characterised by their first and second formants. With high front vowels and r-coloured vowels, it is advisable to measure also the third formant (Ladefoged 2003: 105). The formants of monophthongs are appropriately measured at an interval near the middle of the vowel where the formants are relatively stable (Lehiste & Peterson 1961: 272; Ladefoged 2003: 104). Hall-Lew (2009: 133) used a two-point strategy and measured the formants at midpoint of the steady-state of the vowel or its highest F1 value and towards the end of the vowel.

The formant measurements were carried out in Praat (Boersma & Weenink 2014) using an adapted version of the Vowel Capture Script (Kendall 2009). The script measured the first three formants at 20%, 50% and 80% from the onset of the vowel. The midpoint values were used for the processing of the monophthongs, the 20% (onset) and 80% (glide) values for the processing of the diphthong.

The goal was to collect at least ten tokens per lexical set and speaker across all speech styles although a desired number of tokens would start at 30 to take intra-speaker variation into account (Da Silva 2007: 129). Ten tokens are far from ideal, but it is accepted to obtain statistically sound results. What is more, it keeps the project manageable for a single researcher. Studies have shown that due to the lack of occurrences, researchers have had to work with even less than ten tokens and have received reliable data nevertheless (Bekker 2008; Hall-Lew 2009; Evanini & Huang 2013). A prerequisite of working with relatively few tokens is that outlier values are excluded from the analysis (Thomas 2011: 159). The cleaning of the data included the generation of a boxplot[7] for the frequencies of F1 and F2 for each speaker and each lexical set to check for outliers (Thomas 2011: 50). These boxplots were counter-checked with the vowel space for each speaker. The formant values of outliers were manually re-measured in Praat. If the new value still seemed to be out of place, the token was discarded.

Another decisive factor is the bandwidth of the formants. The lower the amplitude of a sound wave, the larger the bandwidth. Khabanyane (1991: 13) and Ladefoged (2003: 117) therefore suggest to discard formants with a bandwidth greater than 400 Hz because the formant tracker may not be able to find the vowel target. This recommendation was considered

[7] The structure of a boxplot is explained in Section 4.7.2.3.

but not categorically followed. Formants with very high bandwidths were immediately discarded, but bandwidths slightly higher than 400 Hz were checked for their formants. If the values fitted into the pattern of the phoneme, it was accepted.

Formant values correspond with the size of the vocal tract. A direct comparison of speakers, especially when they differ in sex and age, is therefore not meaningful (Thomas 2011: 160). In order to eliminate anatomical differences between the speakers, the formant values were normalised. Vowel normalisation methods distinguishes two general groups – vowel-intrinsic and vowel-extrinsic. Vowel-intrinsic methods use information of a single vowel. Vowel-extrinsic methods compare formant values of different vowels (ideally, but not necessarily, of the whole vowel inventory). Within these methods, it is possible to apply speaker-intrinsic or speaker-extrinsic normalisation. Speaker-intrinsic methods consider the vowel values of a single speaker whereas speaker-extrinsic methods draw on information of more than one speaker. The nature of the present dataset required a vowel-extrinsic and speaker-intrinsic normalisation method. The three most efficient and robust techniques in this respect are the Lobanov method, the Nearey method and the Watt and Fabricius method[8]. The present study used the Lobanov normalisation method (Thomas & Kendall 2007–2015), but the other two techniques could have been equally effectively applied. Vowel normalisation and plotting was carried out with the statistical computing software R (R Development Core Team 2013).

Vowel-extrinsic methods provide more reliable data the more phonemes are used. Phonemes on the margins of the vowel space are particularly important. Therefore, although the present analysis focussed on monophthongs only, the diphthong GOAT was included in the data for vowel normalisation. The onset of GOAT may be [o] and would thus be a potential carrier of the backest sound in the system.

4.7 Statistical analysis

4.7.1 Token selection and model design

The quality of a vowel is influenced by its phonological environment as has been shown in various studies (e.g. Ladefoged 2003; Labov et al. 2006; Sharbawi 2006; Mesthrie 2010b). Nevertheless, the normalised output of Praat, that is, the dataset with all phonological contexts – suitable as well as unsuitable for further analysis – was used for the descriptive statistics

[8] A comparison of these normalisation methods can be found, for example, in Thomas (2011) and Thomas & Kendall (2007–2015). See also Adank et al. (2004).

of vowel quality. The reason for this somewhat unconventional decision was to obtain a general picture of the BSAE vowel space including all phonological environments that can be found in the dataset. For the descriptive statistics of vowel duration and the statistical modelling, however, the dataset was further manipulated to ensure accuracy of measurement. For example, Di Paolo et al. (2011: 88) suggest to treat some contexts, at least initially, as separate categories:

- preceding or following liquids or clusters (e.g. *fleece, dress, strut, hill, bard*)
- following nasals (e.g. *bin, ham, hang*)
- following velars in the case of /æ/ (e.g. *hag, hack*).

Furthermore, Thomas (2011: 49) reports the influence of coronal consonants on the production of neighbouring vowels: Coronals raise the F2 of back vowels, i.e., they are produced more fronted, and they lower the F2 of front vowels, i.e., they are produced more retracted. The data for the regression analysis was thus prepared as follows: Preceding and following /l/, /r/, /w/ and following nasals were excluded. The pre-vocalic context was then divided into the categories:

- preceding coronals: /t d s z θ ð ʃ ʒ n tʃ dʒ/
- preceding non-coronals: /p b f v k g h m/ and
- preceding /j/ for the comparison within the lexical set GOOSE.

Function words were generally excluded because of possible vowel reduction (e.g. Hoffmann 2011: 152). Exceptions were made when the function word was emphasised (cf. Di Paolo et al. 2011: 88) as was frequently the case with *this* and *so* in the Reading Passage:

> ***This** gave the boy so much pleasure […].*
>
> *And **so**, the wolf had a feast.*

The duration of a vowel can have an impact its quality as well. Therefore, vowel sounds shorter than 50 milliseconds were excluded from the statistical analysis because they are prone to centralisation (Evanini & Huang 2013: 5; Hall-Lew 2009: 132–133). Vowel sounds longer than 300 ms were also excluded since they derived from utterances with drawling intonation. For the statistic modelling of vowel duration, tokens in word-final position (e.g. *see, prefer, bamboo*) followed by a pause were omitted because they tend to be pronounced generally longer than tokens in non-word-final position. Apart from that, tokens in word-final position occur exclusively in tense vowel and have no lax correspondent.

Vowel duration can be influenced by voicing and manner of the following consonant. In L1 varieties, vowels are produced shorter when followed by a voiceless consonant and longer when followed by a voiced consonant. Vowels are shortest before voiceless plosives and longest before voiced fricatives. Nasals and voiced plosives have nearly the same influence on vowel duration and fall between voiceless plosives and voiced fricatives. Affricates and plosives influence the duration of the preceding vowel in the same way. The nature of prevocalic consonants appears to be negligible (Peterson & Lehiste 1960: 702).

Considering the above recommendations and findings, statistical models were developed for each vowel cluster. The dependent variables were formant F1 (vowel height), formant F2 (vowel frontness) and vowel duration. Regarding vowel quality, the repertoire of fixed factors and their variants were

- *phoneme* (two or more phonemes)
- *phonological context*
    - *preceding context* (coronals, non-coronals, and J-words for the analysis of GOOSE)
    - *adjacent context* (high front and centralised for the analysis of KIT),
- *spelling* (for the analysis of NURSE and STRUT)
- *speech style* (IN, ReP, WL)
- *gender* (male, female)
- *age group* (young, middle, older)
- *number of vowels in L1* (5, 7)

and the interactions of

- *phoneme*gender*
- *phoneme*age group*
- *age group*gender*
- *preceding context*gender* (for the analysis of GOOSE) and
- *preceding context*age group* (for the analysis of GOOSE).

Post-vocalic contexts were controlled for in that only obstruents entered the models. Sonorants (/l, m, n, ŋ, ɹ, w, j/) can influence vowel quality, may be difficult to distinguish from vowels in the spectrogram and were thus omitted.

Regarding vowel quantity, i.e. vowel duration, the variables differed somewhat from those of vowel quality. *Phoneme, speech style, gender, age group* and *preceding context* and the above interactions remained the same. The variable *number of vowels in L1* was irrelevant for this part of analysis and was thus neglected. Since voicing and manner of the postvocalic consonant plays a role, the variables *following voicing* and *following manner* was included. The fixed factors and their variants for vowel duration were:

- *phoneme* (two or more phonemes)
- *phonological context*
  - *preceding context* (coronals, non-coronals, J-words for the analysis of GOOSE)
  - *following voicing* (voiced, voiceless, pause/gap)
  - *following manner* (fricative, plosive, affricate, pause/gap)
- *spelling* (for the analysis of FOOT)
- *speech style* (IN, ReP, WL)
- *gender* (male, female)
- *age group* (young, middle, older)

and the interactions of

- *phoneme*gender*,
- *phoneme*age group*,
- *age group*gender* and
- *phoneme*speech style*.

Random factors were always *speaker name* (all speakers) and *word label.* Not all variables were analysed for every vowel cluster but instead selected according to the relevance and former findings for the cluster. Deviations from this modelling procedure will be explained in the respective section. The models were designed in a step-down manner: at first, all fixed effect variables and interactions entered the model. After the first model run, all interactions that did not show statistical significance were excluded. After a second model run, all single variables that did not show statistical significance were excluded. In a third model run, usually only statistically significant variables and interactions remained.

To distinguish the dataset for the descriptive statistics and that for the statistical models, the former contains the normalised "raw" data (all possible phonological contexts), and the latter contains the normalised "cleaned" data (only phonological contexts).

4.7.2 Descriptive statistics

4.7.2.1 Analysis of vowel quality

The distance between two vowel categories can be measured in different ways. Nycz & Hall-Lew (2014) compared the following four methods in the measurement of vowel mergers in three L1 English varieties:

1. measurement of the Euclidean distance (see also Labov et al. 2006)
2. mixed effects regression & adjusted Euclidean distance (Nycz 2013)
3. calculation of Pillai-Bartlett trace (Hay et al. 2006; Hall-Lew 2009)
4. determination of the spectral overlap (see alo Wassink 1999).

As all four have advantages and disadvantages, Nycz & Hall-Lew (2014) suggest the use of at least two different methods. Of the above-mentioned approaches, the present study employed the Pillai-Bartlett trace (henceforth Pillai score), which was used, for example, by Hay et al. (2009) for the investigation of the diphthongs NEAR and SQUARE in New Zealand English. Strictly speaking, it is a test statistic that does not directly calculate actual distances or the extent of overlap of two vowels but determines the "abstracted *difference*" of two vowel categories (Nycz & Hall-Lew 2014: 5).

The Pillai score is a result of MANOVA (multivariate analysis of variance). It is a parametric statistic that makes assumptions about the data, "asking what proportion of the total variability in the data is 'explained' by the difference (in means) between the two categories" (Johnson 2017). Pillai scores range from 0 to 1. The lower the Pillai score, the smaller the distance between the formants of two vowels and thus the smaller the phonemic distinction. Reversely, a Pillai score towards 1 indicate two distinct vowels. MANOVA also provides a *p* value for each Pillai score that specifies whether the difference between vowels is statistically significant or not (Nycz & Hall-Lew 2014: 5). A disadvantage of this method is that even if two vowel distributions are fully distinct, there is still residual variation in each distribution so that the Pillai score hardly reaches 1. Similarly, if the means of two distributions are equal, the Pillai score will come out as 0 even if the distribution have different shapes and do not show complete overlap. Another disadvantage is that if the clusters have different token numbers, the Pillai score will go down, thus indicating more overlap than there possibly is (Johnson 2017). This is illustrated in Figure 4.4 and Figure 4.5. The first graph shows two distributions with an identical token number of N = 500. The Pillai score is 0.58.

The distributions in the second graph have the same settings except for the token number of the blue one which is reduced to N = 190. The Pillai score goes down to 0.47 suggesting a higher degree of overlap although this is not the case.

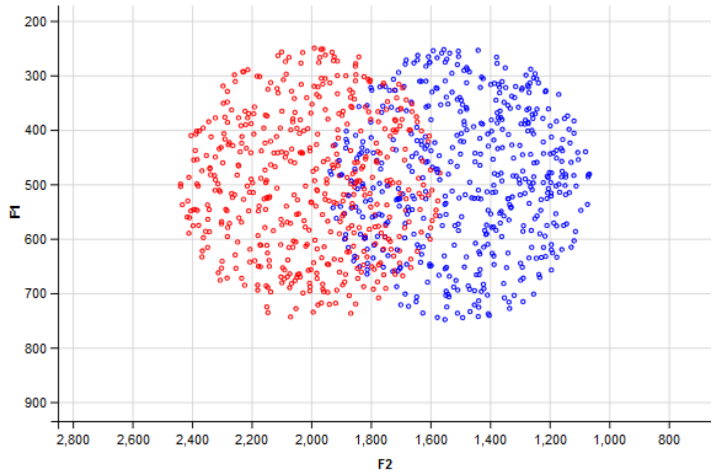

Figure 4.4: Two distributions with equal token numbers (adapted from Johnson 2014)

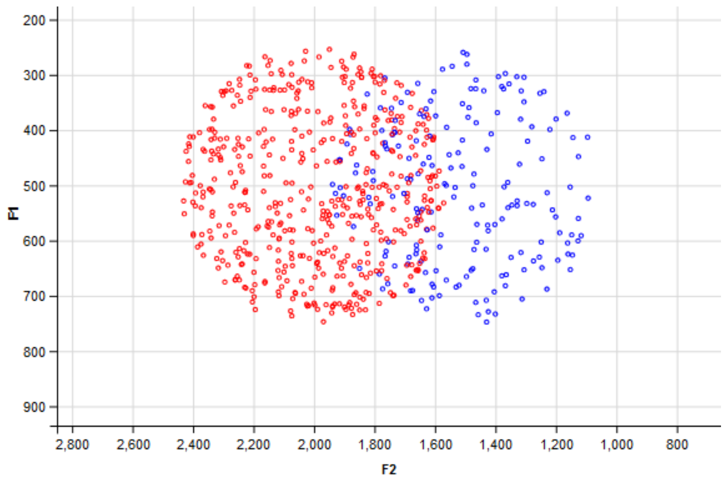

Figure 4.5: Two distributions with different token numbers (adapted from Johnson 2014)

The second method employed is the calculation of the Bhattacharyya coefficient (henceforth BC), which measures the overlap of two continuous distributions, e.g. the tokens of two vowel categories (Johnson 2017; Strelluf 2016). BC also ranges from 0 to 1, but in contrast to Pillai, 0 indicates complete phonemic distinction, and 1 complete overlap. A big advantage is that it mainly avoids the problems explained for Pillai: It reaches 0 if the tokens of two classes clearly do not overlap. It is "appropriately less than 1" if the distributions have the same mean, but technically do not overlap (e.g. when the distributions are skewed or if distributions have the same centre but different angles). And finally, it is less sensitive to imbalanced token numbers (Johnson 2017). A disadvantage, however, is that the calculation of the BC is not a test statistic. Hence, it does not provide significance values.

Both calculations, Pillai score and BC, were used because the advantages of one method were expected to compensate for the disadvantages of the other, and the results of both would provide a balanced picture. The calculations were carried out for each speaker and each meaningful vowel pair, for example FLEECE vs. KIT, NURSE vs. DRESS, BATH vs. STRUT etc. The computation platform for both calculation methods was the statistical computing software R (R Development Core Team 2013). The BC was calculated with the *adehabitatHR* package (Calenge 2006) and the *sp* package (Pebesma & Bivand 2005). The calculation of the Pillai scores did not require additional R packages.

4.7.2.2 Analysis of vowel duration

Languages with contrastive vowel length distinguish between phonologically short (lax) and long (tense) vowels. The ratio of English tense/lax vowels is 1.2:1, which means that a short vowel is on average 83% of the length of a long vowel. If the vowel precedes a voiced obstruent, the ratio increases to 1.5:1, meaning that a short vowel is on average 67% of the length of a long one (Wassink 2006: 2335; see also Thomas 2011: 143). The voicing of a preceding consonant does not seem to affect vowel length (Peterson & Lehiste 1960: 702). While Germanic languages utilise contrastive vowel length, second language varieties of English may not. Since vowel length is not phonemic in Bantu languages, lax and tense vowels of L2 English may be randomly produced. Three examples from the present dataset shall illustrate this phenomenon:

- FLEECE longer than KIT: *seat* 149.7 ms, *sit* 69 ms (speaker SMM10)
- KIT longer than FLEECE: *seat* 84.8 ms, *sit* 102.8 ms (speaker OYF13)
- both vowels of same length: *seat* 67,6 ms, *sit* 64.8 ms (speaker SMM13).

The absolute duration of a sound depends, among others, on the speed of the utterance. An average rate of delivery may contain anything between about 6 to 20 sounds per second (Thomas 2011: 192). In order to examine contrastive and contextual vowel length and to compare speakers with different speech rate, vowel duration was normalised with a three-step method adapted from Wassink (2006: 2345). The first step is to calculate the mean duration for each phoneme category (i.e. individually for FLEECE, KIT, DRESS, etc.) for each speaker. The second step is the calculation of the grand mean $\overline{D}_k$ per speaker:

$$\overline{D}_k = \frac{\sum_{j=1}^{n}\left(\sum_{i=1}^{n} D_{ijk}/n_j\right)}{n}$$

The grand mean of duration $\overline{D}_k$ for speaker k is calculated across all phoneme category durations where D_{ijk} is the observed duration for token i of phoneme category j for speaker k.

The final step calculates the normalised duration for each vowel token. The normalised duration δ_{ijk} is equal to the duration of the individual token minus the grand mean:

$$\delta_{ijk} = D_{ijk} - \overline{D}_k$$

Since vowel length is assumed to be conditioned also by speech style, the calculations were carried out separately for each style. The normalised duration of the individual tokens can assume positive and negative values. Positive values indicate that the vowels are longer than the grand mean, and a negative value means that the vowel is shorter than the grand mean.

Table 4.4 shows the example of FLEECE and KIT in the speech style IN for two participants. Both speakers have a similar grand mean across all phonemes, and they have identical values for KIT, which means that they produce KIT 21.5 ms shorter than their respective grand mean. The speakers differ, however, in their normalised length for FLEECE: Speaker DOF06 produces FLEECE 17.6 ms longer than her grand mean and 39.1 ms longer than KIT. Speaker NMF02 produces FLEECE 16.3 ms shorter than her grand mean and only 5.2 ms longer than KIT. Thus, both speakers differ in their lax/tense distinction.

Table 4.4: Normalised duration (ms) for FLEECE and KIT of two speakers in Interview style (IN)

Speaker	Grand mean across all phonemes per speaker	Normalised duration for KIT	Normalised duration for FLEECE
DOF06	100.1	-21.5	17.6
NMF02	101.9	-21.5	-16.3

4.7.2.3 Graphic presentation of the results

Vowel quality will be graphically presented with vowel plots and contour plots. Figure 4.6 shows the vowel plot of the Lobanov-normalised mean values for FLEECE and LOT. The Lobanov method transforms the Hz values of the vowel formants into standard deviations (SD) from the mean of the most central vowel in a speaker's vowel space, denoted by the intersection of the zero points (Thomas & Kendall 2007–2015). In the vowel chart, the mean formant values for FLEECE are F1 = -1.16 and F2 = 1.51. That means that the position of FLEECE in the vowel space is 1.16 SD higher and 1.51 SD fronter than the midpoint. By contrast, the position of LOT is F1 = 0.46, F2 = -0.97 and thus 0.46 SD lower and 0.97 SD backer than the midpoint.

For a better illustration of the token distribution, standard deviations (1 SD) were added. The SDs are shown as dashed circles and comprise 68% of the tokens of each phoneme. In order to avoid confusion with a twofold use of SD (on the one hand as the units of measurement of the Lobanov transformation, and on the other hand as the measure of dispersion in the dataset), the distance of a vowel from the midpoint will be called Lobanov Unit (LU).

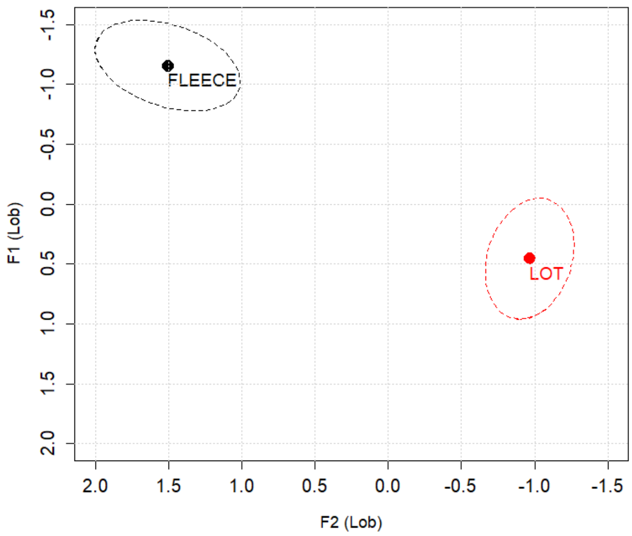

Figure 4.6: Vowel chart of FLEECE and LOT

Vowel plots provides a good overview of vowel distribution and overlap. Sometimes, however, a more fine-grained depiction of vowel overlap is necessary. In these cases, contour plots will be employed as seen in Figure 4.7. A contour plot is calculated by means of kernel density estimation (KDE), "which is a non-parametric way to estimate the probability density function of a random variable" (DiCanio 2013). Each contour line represents points in the F2–F1-plane with a certain constant probability density value. For points which do not lie on such a contour line, the corresponding probability density value can only be estimated. The value always lies in the interval given by the two neighbouring contour lines. Here, the outer contour lines are associated with lower values of the probability density than the inner contour lines. Consequently, the small contour lines in the middle of the figure stand for the area in the F2–F1-plane where the probability density function is the highest. Contour plots are not symmetric (compared to the ellipsis of the SDs in Figure 4.6) but describe the actual spread of tokens and thus provide "a clearer sense of the concentration of observations" (DiCanio 2013). Still, this depiction will be the exception because in clusters with three or more vowels, a contour plot can be messy. In these cases, the vowel plot with SDs is a better choice.

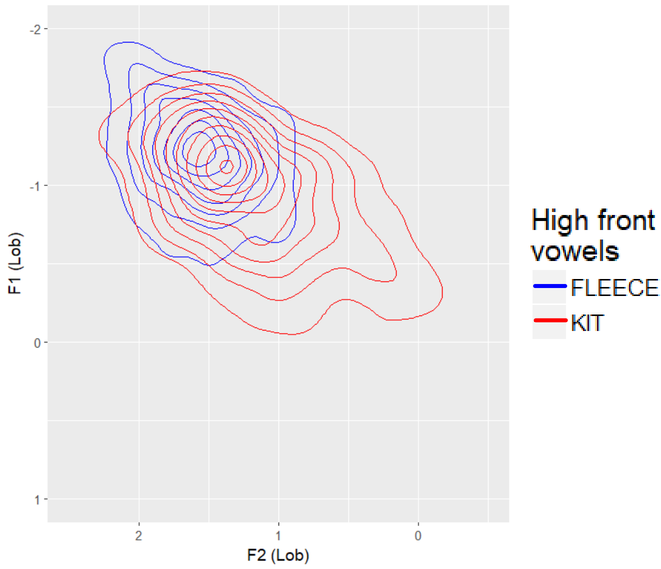

Figure 4.7: Contour plot of FLEECE and KIT

Vowel duration will be illustrated by means of boxplots (Figure 4.8). A boxplot is a graphic representation of numerical data by their quartiles. The box itself contains the middle 50% of the observations (the interquartile range, IQR). The band inside the box is the median, the middle value of a set of ordered observations. The dashed lines are the so-called whiskers, which denote the lower and upper quartile. Their ends represent the lowest and highest scores and are not longer than 1.5 IQR of the lower and upper quartile respectively. The individual dots outside the plot are the outliers of the dataset (Field et al. 2012: 914; Levshina 2015: 58).

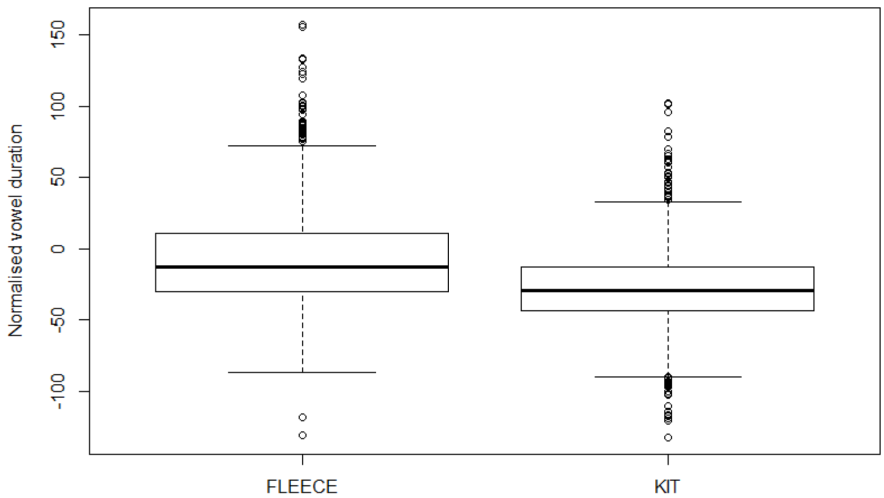

Figure 4.8: Boxplot of FLEECE and KIT

4.7.3 Statistical modelling

The computation platform for the statistical analysis was the computing software R (R Development Core Team 2013). The analysis was carried out by multiple linear mixed effects regression (lmer) models using the *lme4* and *lmerTest* packages for R (Bates et al. 2015; Bates et al. 2017; Kuznetsova et al. 2017a; Kuznetsova et al. 2017b) for each vowel cluster. Multiple linear regression is a statistical procedure that models the relationship between a dependent variable and two or more independent variables (Levshina 2015: 121). Thus, it is possible "to measure the individual impact of each variable in the model while controlling for the other variables" (Levshina 2015: 121). The attribute "linear" means that the relationship between both variable types is proportional, i.e. the mean values of the dependent variable for each increment of the independent variables lie along a straight line (regression line) in a graph (Field et al. 2012: 246). Apart from analysing individual variables, it is also possible to model interactions between independent variables in order to detect the combined influence of them, for example the effect of *gender* in combination with *age group* on vowel height (F1).

A mixed model uses two types of factors[9] which are responsible for the outcome – fixed factors and random factors (e.g. Bates et al. 2015: 1; Johnson 2009: 364; Winter 2013: 22). Fixed factors have a small and exhaustive number of possible variants or levels, for example *gender:* male/female, *age group:* young/middle/old, *phonological context:* obstruent/sonorant), which can be replicated in future studies. They are the objects of interest and expected to have a systematic and predictable influence on the data (Johnson 2009: 365; Winter 2013: 39). For fixed effects, the mixed model produces coefficients associated with the differences between the variants.

Random factors are drawn from a larger population, for example *speaker* and *word*. They are usually not replicable since future or parallel studies may not involve the same speakers or the same words. Random factors are expected to have a non-systematic and unpredictable influence on the data (Johnson 2009: 365; Winter 2013: 39). For them, the model estimates a single parameter that represents the degree of inter-speaker variation (Johnson 2009: 365). The advantage of random factors in a model is explained by Brato (2016: 79), who investigated pronunciation differences among adolescents in Aberdeen, Scotland: "Including speaker as a random effect means that if an individual's behaviour deviates from the estimate for the rest of 'their' group – be it males, middle-class speakers or ethnicity –, this behaviour can still be taken into account." He continues:

[9] synonymous for the terms 'variables' and 'effects'

> A mixed model still captures other external factors, but only if they can contribute to the explanatory power more than the inter-speaker variation. Thus, whereas a fixed-effects-only model will often include quite a large range of factors, making individual effects rather difficult to interpret, a mixed-effects model can clearly reduce the number of significant factors and simplify their interpretation (Brato 2016: 79).

Therefore, such a model "is particularly useful in a situation [...] in which language change is promoted or inhibited by individual speakers" (Brato 2016: 79). Since this scenario may also apply for the data under investigation, a mixed effects model was used in the present study.

The output of the statistical analysis is based on REML t-tests using Satterthwaite approximations to degrees of freedom. The regular lme4 package for R does not provide p-values for the fixed effects "because there is some question about what the correct degrees of freedom for the t-test should be" (Johnson, p.c. 2019). The lmerTest package, however, estimates the degrees of freedom via Satterthwaite approximations and thus provides p-values and significance levels for the fixed effects (Kuznetsova et al. 2017b). REML (restricted (or residual) maximum likelihood) is the part of the statistic that estimates the variances of the random effects (Oehlert 2011: 4). The REML method can handle equal and unequal sample sizes as well as unequal variances (O'Neill 2013). The coefficients necessary for the interpretation of the results will be explained in the following. The made-up example in Table 4.5 is used for illustrating the output design of an lmer model for F2 (vowel frontness) for KIT and FLEECE.

Table 4.5: Made-up example of an lmer output for F2 (LU) in KIT and FLEECE

```
Scaled residuals:
                     Min        1Q     Median      3Q       Max
                  -6.0261   -0.5238   0.0583   0.6249   2.7766

Random effects:
 Groups           Variance   Std.Dev.
 word label       0.0406     0.2015
 speaker name     0.0051     0.0717
 Residual         0.1983     0.4453

Fixed effects:
                    Estimate  Std. Error    df     t value    Pr(>|t|)
 (Intercept)          1.5230     0.0391   182.10    38.974     <2e-16 ***
 ph kit              -0.4208     0.0457   169.10    -9.217     <2e-16 ***
 gender male          0.0569     0.0321    38.90     1.772     0.0842
 style ReP            0.0059     0.0583   473.30     0.101      0.919
 style WL             0.1275     0.0518   175.00     2.459      0.015 *

R²m (fixed effects) = 0.16
R²c (fixed + random effects) = 0.32

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

Residuals are the deviations of the observed data points from the predicted values. The output describes the distribution of the residuals without making statistical assumptions. It shows the values for the median (the middle value of a set of ordered observations), the first quartile (1Q), the third quartile (3Q) and the end points (Min and Max) of the distribution. Residuals can take positive values when they occur above the regression line and negative values when they occur below it (Levshina 2015: 120; Winter 2013: 12). They should be normally distributed and centre around zero. In large datasets, this assumption becomes less important though (Levshina, 2015: 155).

Random effects show the estimated variance and standard deviation of those independent variables that cannot be predicted. In the present study, these are the variables *speaker name* and *word label*. Furthermore, they show the estimated variance and standard deviation of the mean residual (Levshina 2015: 120; Winter 2013: 12).

Fixed effects show the values for the independent variables that can be predicted. The first column (Estimate) of the fixed effects provides the estimated mean of the intercept and the slopes of the regression line. The output table also specifies the standard errors (estimate of precision) of the estimated coefficients and covariate t-values for testing the null hypothesis. The t-value is the quotient of the estimate and the standard error. Finally, it shows the probability value p, which is based on the t-statistics (Levshina 2015: 145; Winter 2013: 4–5). The fixed factors of the model and their variants were *gender* (male/female), *speech style* (IN/ReP/WL) and *phoneme* (FLEECE/KIT). Whereas the estimates for male, ReP, WL and KIT are given explicitly in the resulting table, the values for female, IN and FLEECE are summarised in the intercept[10]. The intercept includes that variant of a factor that comes first in the alphabet: (female comes before male, FLEECE comes before KIT and IN before ReP and WL). The variants in the intercept form the reference levels to which the other variants of each factor are compared. Hence, the row labelled 'gender male' compares the male speakers with the value of the females hidden in the intercept, and so on. The Intercept estimate shows that the F2 mean for female speakers for the phoneme FLEECE in the style IN is 1.523. The estimated F2 mean value for KIT is -0.4208, i.e., KIT tokens were produced on average 0.42 LU more retracted than FLEECE. The corresponding p-value is lower than 0.001 and indicates a statistically significant difference between both phonemes. The estimated F2 mean for male speakers is

[10] Mathematically, the intercept is the value at which the regression line crosses the y-axis (Baayen (2008: 85). This value itself is usually meaningless and thus not worth interpreting, but it must be employed to obtain physically meaningful data of the sample under investigation (Frost 2013).

0.0569. This means that males have an F2 value which is about 0.06 Lobanov Units (LU) higher than that of females. In other words, males produce F2 slightly more fronted than females. The p-value of 0.08 shows that this outcome is statistically not significant. The parameter *speech style* is shown twice, as style ReP and style WL. Their corresponding estimates show F2 in comparison to the reference variant style IN. The estimate for ReP is 0.0059, which means that words in the reading passage were produced with practically the same F2 as the words in the interviews. The value for WL is 0.1275 higher than that of IN, i.e., words in the wordlist were produced about 0.13 LU fronter than words in interview style. WL has a p-value of 0.015, thus, the difference in this speech style is statistically significant.

Another important parameter is $\mathbf{R^2}$ (R squared), the coefficient of determination, which is a measure of the explanatory power of the model. It reflects how much of the outcome is explained by the independent variables in the model and ranges from 0 (0% explained) to 1 (100% explained) (Field et al. 2012: 250; Rasinger 2013: 171). R^2 is subdivided into $\mathbf{R^2m}$ (marginal R squared) and $\mathbf{R^2c}$ (conditional R squared). R^2m refers to the fixed effects, and R^2c to the fixed effects plus random effects. Generally, the higher R^2 the better the quality of the model. However, a low value does not necessarily hint at a flawed dataset (Frost 2014). In L2 varieties, pronunciation variability and thus dispersion may be higher than in L1 varieties. Therefore, a low R^2 can still come along with meaningful results. The value for R^2c in this example is 0.32 and thus moderately high. The fixed factors (16%) and the random factors (16) account for 32% of the variability in F2 (vowel frontness).

For the post-hoc analysis of single variables, Tukey's HSD test was employed. It controls Type I error rate (the identification of a chance effect as a real one) very well. It is less reliable with Type II errors though, i.e. the failure to identify an effect that actually exists is high (Field et al. 2012: 431). However, it is probably more beneficial to take "a conservative approach, arguing that it is better to overlook something that does exist than to report something that does not" (Johnson 2009: 369). The post-hoc tests of interactions were carried out with the Tukey test as pairwise comparisons via least square means (lsmeans). The calculation of least square means is particularly recommendable when the observations for each combination is not equal (Mangiafico 2016a, 2016b). This is the case for the present study. The significance level used throughout the analyses was $\alpha = 0.05$.

5 Results and Discussion

5.1 Questionnaire data

All speakers filled in a questionnaire, the design of which can be found in Appendix A-2. Section 4 of the questionnaire inquired about the language skills and habits of the participants. The first question asked for the language the participants first acquired. These languages were in ascending order Southern Sotho 21, Xhosa 8, Tswana 7, Zulu 5, Ndebele 1, Northern Sotho 1, Venda 1 (see also Table 4.1). Since South Africa is a multilingual country, it can be assumed that the language repertoire of the participants not only includes their mother tongue and English, but also other languages. Therefore, the following two questions were included:

(4.2) What other South African languages did you acquire?

The number of additionally spoken languages can be seen in Table 5.1. Forty participants spoke two or more languages. The four participants who stated only one additional language, ticked English. This result is a vivid example of the bi- and multi-lingual repertoires in South Africa as described by Coetzee-Van Rooy (2014: 39). The number of participants who speaks a particular South African language is shown in Table 5.2.

Table 5.1: Number of languages spoken additionally to the language first acquired (N = 44)

Additional language(s)	1	2	3	4	5	6	7	8	9	10
Speakers	4	5	12	5	7	5	3	1	1	1

Table 5.2: Number of participants speaking a particular South African language (N = 44)

Additional language	Afrikaans	Zulu	Xhosa	Tswana	N. Sotho	S. Sotho	Tsonga	Swati	Venda	Ndebele	Shona	Chichewa
Speakers	29	24	19	18	13	12	7	6	2	1	1	1

English was not considered since all of them spoke it as a second language. Of the languages mentioned, Afrikaans was spoken by most participants. Considering the age groups, Afrikaans was chosen by all older participants, 11 (out of 17) middle-aged and 11 (out of 20) young participants. This outcome may be due to the “environmental effect” which refers to the selection of languages that are widely spoken in the immediate environment (Coetzee-Van Rooy 2012: 95). Since Afrikaans is a major language in the region, it is not surprising that it is used by many people. What is more, in the Vaal Triangle, many job opportunities are available especially for speakers of Afrikaans, so it is beneficial to speak it (Coetzee-Van Rooy 2012: 112). Another reason may be that at least some participants learnt Afrikaans at school and thus have command of the language.

(4.3) Which language do you speak best?

This question investigated the perceived language strengths. Most informants (39) indicated that the language they first acquired is the one or one of those they spoke best. Of them,

- 17 speakers chose only the language first acquired as their strongest one,
- one speaker chose the language first acquired + a language within the same language family,
- seven speakers chose the language first acquired + English (of these, one speaker also chose Afrikaans),
- one speaker chose the language first acquired + Afrikaans,
- 13 speakers chose the language first acquired + one or more indigenous languages + English (of these, four speakers also chose Afrikaans).

Five speakers chose other languages than those first acquired of which three speakers (XYF03, ZYM06 and ZYM09) selected English as their language spoken best. In total, 20 speakers opted for English as their strongest or one of their strongest languages. The overall results correspond with the findings of Coetzee-Van Rooy (2012: 95). In her study of the language behaviour of Southern Sotho and Zulu speakers, she found that the majority of her respondents perceived their home language as their strongest language (Coetzee-Van Rooy 2012: 95). However, about one-third of her participants stated English as their strongest language. Despite the large number of these respondents, Coetzee-Van Rooy does not interpret this finding as an indicator of a potential language shift (cf. Deumert 2010). On the contrary, she rather sees evidence for "functional multilingualism", i.e. home languages are maintained in the family domain, and English is used in education and in the media (Coetzee-Van Rooy 2012: 111). Furthermore, she reports that the multilingual repertoire of young speakers is expanding because it also includes Bantu languages beside English (Coetzee-Van Rooy 2012: 87). Coetzee-Van Rooy (2012) also observed the "family effect", which refers to the acquisition and use of languages that are related to one's home or strongest language. The present data does not show this outcome. Only one speaker remained within the language family of the language she first acquired. Twenty-one speakers indicated languages within as well as outside their language family. The present findings cannot be directly compared to the results of Coetzee-Van Rooy's (2012) study because of the different wordings of the questions. However, in conclusion it can be said that multilingualism is "flourishing" (Coetzee-Van Rooy 2012: 87) among the participants in the present study, and within this multilingualism, English plays an important role.

(4.4) How often do you speak English in the following context?

This question elicits the perceived frequency with which the participants speak English in certain contexts (Table 5.3). Participants who were university staff members answered only the category 'at work' (not 'at university'). Students answered the category 'at university'. Seven young participants stated that they speak English at work; they were either student assistants or had part-time jobs.

In the private sector, the participants use English, but not as often as in the professional domains. In conversations within the family, most young and middle-aged participants use English sometimes or frequently. The older participants who answered this question are split: three use English sometimes or frequently and three practically never. Deumert (2010) investigated the language shift from African languages to English and Afrikaans in urban areas by means of census data. She compared the years 1996 and 2001 and reported a trend of multilingualism including English in the homes of black South Africans. Of course, this conclusion cannot be drawn from the present data, but one outcome is obvious: non-native speakers of English use English at home.

In conversations with their friends, young participants use English more frequently than other participants. Eighteen young speakers (90%) use English at least sometimes. But also fifteen middle-aged and four older speakers (88% and 57% respectively) use English at least sometimes when talking to their friends. None of the participants stated that they never use English in this context. The category 'at the market/in town' provides a very similar picture. Finally, all participants stated that they frequently or always speak English at work and at the university respectively, and the majority chose 'always'.

As can be seen, English is employed in practically all domains of private and public life to different degrees though. In a multilingual society, however, it is not surprising that English is only one language in the linguistic repertoire of South Africans.

Table 5.3: Number of participants who speak English in the following contexts (N = 44)

	always	frequently	sometimes	rarely	never	no data	not applic.
in your family							
Young	0	6	7	3	3	1	0
Middle	1	4	9	2	1	0	0
Older	0	1	2	2	1	1	0
Total	**1**	**11**	**18**	**7**	**5**	**2**	**0**
with your friends							
Young	6	10	2	1	0	1	0
Middle	5	4	6	1	0	1	0
Older	1	2	1	2	0	1	0
Total	**12**	**16**	**9**	**4**	**0**	**3**	**0**
at the market/in town							
Young	5	6	5	2	1	1	0
Middle	5	1	8	2	0	1	0
Older	1	1	2	2	0	1	0
Total	**11**	**8**	**15**	**6**	**1**	**3**	**0**
at work							
Young	6	1	0	0	0	0	13
Middle	15	2	0	0	0	0	0
Older	3	2	0	0	0	0	2
Total	**24**	**5**	**0**	**0**	**0**	**2**	**13**
at university							
Young	14	5	0	0	0	1	0
Middle	0	0	0	0	0	0	17
Older	0	0	0	0	0	0	7
Total	**14**	**5**	**0**	**0**	**0**	**1**	**24**

Section 5 of the questionnaire elicited the participants' attitudes towards English.

(5.1) In your opinion, which of the following languages is (or are) the key for power and success?

With one exception (a young male), all speakers opted for English. Afrikaans ranks second far behind English (ten votes, 23%), followed by Southern Sotho (20%) and Zulu (18%). The remaining languages obtained less than 10% of votes, four of which none (see Table 5.4). The result is not surprising: Most respondents opted for English because this language is linked with social mobility and progress in life (see also Section 2.1 and 2.2). Bantu languages largely lack this appeal. Interestingly, Afrikaans, providing economic opportunities in the area of the fieldwork (Coetzee-Van Rooy 2012: 92) did not find much approval. Most respondents probably considered this question not in a regional but rather national or even global context and thus chose English as the perceived uncontested language for professional and personal advancement.

Table 5.4: Number of participants who chose the following language(s) as the key for power and success

Language	English	Afrikaans	S. Sotho	Zulu	Twana	Xhosa	N. Sotho	Venda	Ndebele	Swati	Tsonga	Others
Speakers	43	10	9	8	4	2	1	1	0	0	0	0

(5.2) How well do you agree with the following statements?

The answers to this question are shown in Table 5.5. 'I like to speak English' was confirmed by 36 participants, and five rather agreed to this statement. This is the majority of 41 speakers. The statement about the speaker's confidence of using English yielded similar results. Forty-one participants agreed or rather agreed. The importance of being good at English was also largely agreed or rather agreed with. The results for the final statement are in stark contrast to those of the preceding ones: Twenty-five participants agreed and rather agreed that English forms part of their identity. This is still more than 50 per cent, but there were more speakers who rather agreed than fully agreed. What is more, 18 participants rather or completely disagreed. This included about half of the middle-aged and the older speakers, but also one fourth of the young. It seems that many participants regard English a useful skill necessary to move forward in life, but they do not connect it with their own identity. This assumption is backed by Coetzee-Van Rooy (2012: 114) who proposes that black South Africans perceive the importance of English, but simultaneously maintain and cherish traditional culture. Van Rooy (2020: 227) draws a similar conclusion as he states: "English in its various forms is firmly rooted in Africa, although it is not the unqualified carrier of civilization and economic development that it might be conceived of in many circles."

Table 5.5: Answers to the question: How well do you agree with the following statements? (N = 44)

	agree	rather agree	rather disa-gree	disa-gree	no data
I like to speak English					
Young	15	4	1	0	0
Middle	15	1	0	1	0
Older	6	0	0	1	0
Total	**36**	**5**	**1**	**2**	**0**
I feel confident using English					
Young	14	5	1	0	0
Middle	13	3	0	1	0
Older	6	0	0	1	0
Total	**33**	**8**	**1**	**2**	**0**
It is important to be good at English					
Young	17	2	0	0	1
Middle	16	0	0	1	0
Older	7	0	0	0	0
Total	**40**	**2**	**0**	**1**	**1**
English forms part of my identity					
Young	4	10	3	2	1
Middle	4	4	4	5	0
Older	2	1	1	3	0
Total	**10**	**15**	**8**	**10**	**1**

The analysis of the questionnaire revealed no major differences between the age groups. Since due to this outcome, statistically significant differences were not expected. Therefore, these results did not enter the statistical models.

5.2 The vowel space of BSAE

This section describes the vowel realisation of the participants. Figure 5.1 shows the Lobanov-normalised "raw"[11] vowel space of BSAE across all participants and speech styles with 44 speakers, 13,919 monophthong vowel tokens and 818 GOAT tokens (N = 14,737).

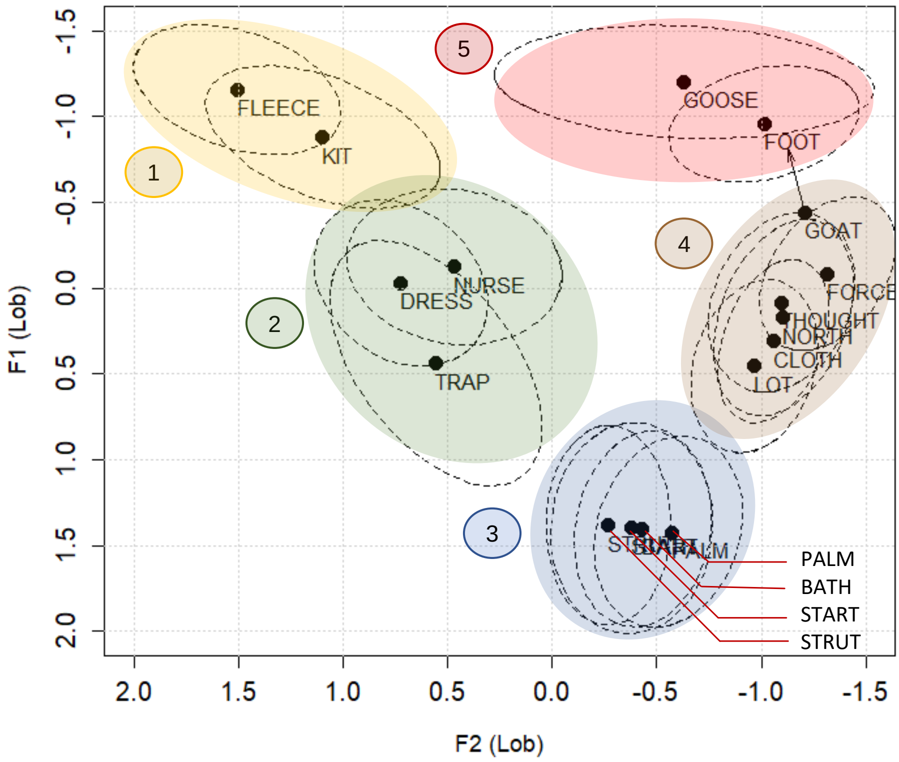

Figure 5.1: Lobanov-normalised "raw" vowel space of BSAE across all speakers and variables with 1 SD (N = 14,737)

The vowel space shows five major clusters:

1. high front: FLEECE, KIT
2. mid front: DRESS, NURSE, TRAP
3. low central: STRUT, START, BATH, PALM
4. mid back: LOT, CLOTH, THOUGH, NORTH, FORCE, GOAT
5. high back: GOOSE, FOOT.

As mentioned in Section 4.6.7, the diphthong GOAT was utilised in the normalisation to include every phoneme that is potentially the backest in the vowel space. Since the focus is

[11] This dataset includes all phonological contexts available even those that were omitted in the statistical modelling because of possible influences on the vowel quality (see Section 4.7.1).

on monophthongs, GOAT will not be considered in further investigations (although its pronunciation may be a monophthong rather than a diphthong in BSAE). In this dataset, FORCE is the backest vowel in the vowel space.

FLEECE is mainly a tense vowel but also occur in the area of KIT. KIT has a large horizontal but also a slight vertical spread. It ranges from tense [i] to centralised [ɪ̈]. In the mid front cluster, DRESS has the smallest spread. NURSE vowels have a horizontal distribution and show a large overlap with DRESS. TRAP also overlaps with DRESS tokens, but it is also realised with more open phonemes as [æ] or [a]. Of the open back cluster, BATH, START and STRUT show a large degree of homogeneity. The mean of STRUT is somewhat fronter, but its SD is practically within the SD of START. PALM, the fourth member of this cluster, is produced much backer and lower than the other three. The mid back cluster is not as homogeneous as is described in the literature. It shows a large vertical distribution with LOT the lowest and FORCE the highest and backest vowel. Of the lax vowels, LOT and CLOTH have different mean values, but their overlap is large. LOT is mainly realised as [ɔ], but there are also occurrences in the /ʌ/ area. Of the tense subcluster, THOUGHT and NORTH have practically the same mean and the same SD. By contrast, FORCE, is significantly backer than the former two. Regarding the high back cluster, GOOSE has an extreme horizontal spread. GOOSE is produced significantly fronter and somewhat higher than FOOT. Their SDs overlap in the back half of the graph.

Based on the above vowel chart, Table 5.6 lists the phoneme values transformed into IPA characters to allow for a direct comparison with Van Rooy & Van Huyssteen's (2000) data (see Section 2.5.1). The dots above the characters denote a centralised realisation of the sound.

Table 5.6: Vowel realisation in the "raw" BSAE dataset

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i] > [ɪ]	STRUT	[ä], [ɐ]
KIT	[i], [ɪ] > [ɪ̈]	LOT	[ɔ], [ʌ]
DRESS	[e], [ɛ]	CLOTH	[ɔ]
NURSE	[e], [ɘ]	THOUGHT	[ɔ]
TRAP	[æ]~[ɛ] > [a]	NORTH	[ɔ]
BATH	[ä]	FORCE	[ɔ], [o]
START	[ä]	GOOSE	[ʉ], [u]
PALM	[ɑ]	FOOT	[u]

The following sections will describe and discuss each vowel cluster. As mentioned in Section 4.7.1, the descriptive statistics of vowel quality were carried out with the raw Lobanov-normalised dataset, in which phonological context was not yet controlled. These measurements include the token number, the means of F1 and F2 and the standard deviations for each lexical set. The values for the Pillai scores and the Bhattacharyya coefficients (BC) also derive from the raw data. They were calculated only for the interview data as this kind of data was expected to produce most informal and most natural speech. The dataset for the descriptive statistics of vowel duration and for the regression models are based on a cleaned dataset with defined pre- and post-vocalic contexts. The number of observations in both statistical approaches may therefore differ. The description of the first cluster, high front, will contain comprehensive output tables for the analysis of F1 (vowel height) to show the general procedure. For the sake of readability and brevity, the remaining output tables will show only statistically significant figures. For better illustration, each section depicts the vowel space of the respective cluster. Vowel plots referring to the descriptive statistics contain Lobanov-normalised raw data. Those referring to the statistical models as well as boxplots contain Lobanov-normalised cleaned data.

5.2.1 The high front vowels

5.2.1.1 Vowel quality

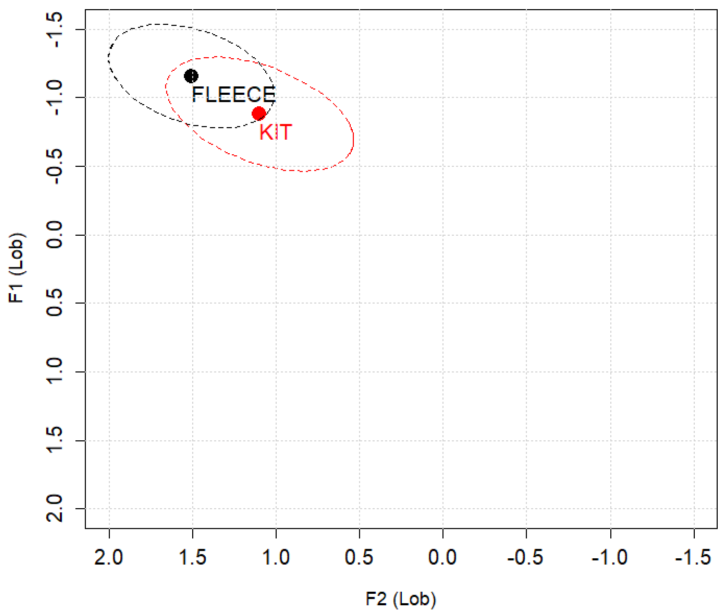

Figure 5.2: High front cluster across all variables with 1 SD

Figure 5.2 shows two significantly distant vowel means but large F2 spreads (KIT: SD = 0.56, FLEECE: SD = 0.50, see Table 5.7). The overlap of the two SDs contains the means of both phonemes. The KIT tokens cover a large area of FLEECE, but some are produced also in centralised position.

Table 5.7: Token numbers, formant means and SDs of the high front cluster

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
FLEECE	1133	-1.16	0.38	1.51	0.50
KIT	1307	-0.88	0.42	1.10	0.56

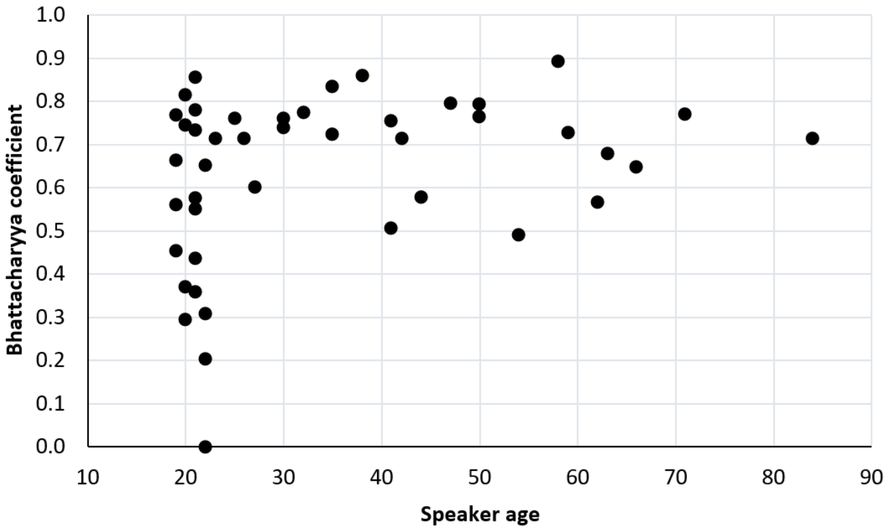

Figure 5.3: Bhattacharyya coefficients for FLEECE (N = 588) and KIT (N = 616) in IN style separated by speaker age

Figure 5.3 illustrates the degree of overlap of FLEECE and KIT tokens in the interview data by means of BC. Many speakers across all ages show a large overlap with BC = 0.87 suggesting that both phonemes are produced with little difference. Pillai = 0.16, $p < 0.001$ confirms the high degree of vowel overlap. Yet, some young speakers show very little or no overlap. These were TYM07 (0.00), who clearly produced two distinct vowels, ZYM09 (0.20), XYF03 (0.30) and SYF10 (0.31). The speaker codes are not shown in the graph.

The regression analysis was carried out for F1 and F2. The fixed factors for both runs were *phoneme, speech style, gender, age group, number of vowels in L1, preceding context* (coronals, non-coronals) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. The variables that yielded statistically significant results for F1 are shown in Table 5.8.

These are the variables *phoneme, gender, age group, preceding context* (coronals, non-coronals) and the interaction of *phoneme*age group*. KIT is produced 0.14 Lobanov Units (LU) lower than FLEECE, However, since the interaction of *phoneme*age group* is also significant, it is advisable to focus on the groups in the interactions rather than to compare the single variables. Furthermore, male speakers produce F1 on average 0.14 LU lower than females. Finally, preceding non-coronals are produced 0.06 LU higher than coronals. Despite the statistical significance, the latter outcome does not seem phonetically relevant. *Speech style, number of vowels in L1* and *phoneme*gender* showed no statistical significance. $R^2c = 0.34$ is moderately high. The outcome can be explained by 16% of the fixed factors and 18% of the random effects. The post-hoc test for the interaction of *phoneme*age group* produced significant differences for the comparisons between middle-young and older-young participants showing that the differences were generated by young speakers. The comparison of middle-aged and older speakers was not statistically significant ($p = 0.70$).

Table 5.8: F1 regression results for FLEECE (N = 766) and KIT (N = 731) (ph = phoneme, ag = age group, pc = preceding context)

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.3418	-0.6405	-0.0678	0.5749	3.7816

Random effects:

Groups	Variance	Std.Dev.
word label	0.0100	0.0999
speaker name	0.0220	0.1420
Residual	0.1138	0.3374

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-1.0227	0.0534	65.40	-19.162	<2e-16	***
ph kit	0.1391	0.0359	229.80	3.872	<0.001	***
ag older	-0.0361	0.0754	51.90	-0.479	0.634	
ag young	-0.1436	0.0547	52.10	-2.625	0.011	*
pc non-coronal	-0.0627	0.0265	138.70	-2.362	0.020	*
gender male	-0.1379	0.0476	40.10	-2.901	0.006	**
ph kit:ag older	-0.0206	0.0533	1424.40	-0.386	0.700	
ph kit:ag young	0.2800	0.0389	1432.50	7.198	9.83e-13	***

R^2m (fixed effects) = 0.16
R^2c (fixed + random effects) = 0.34

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'

The results of the post-hoc test for interactions provide estimates that cannot be straightforwardly detected in a vowel plot. Therefore, a pairwise comparison of the factors was used to calculate the estimates in all combinations. Table 5.9 shows the statistically significant results of the interaction of *phoneme*age group*. Six comparisons out of fifteen combinations are statistically significant. With one exception, the young speakers produce these significances: The KIT and the FLEECE vowel of young speakers differs from KIT and FLEECE of the other age groups. The phonemes are produced differently within the young and the middle-aged speakers, but not within the older speakers. The young produced the largest distance between KIT and FLEECE, that is, they pronounce FLEECE 0.42 LU higher than KIT. Interestingly, there is no statistical significance between the age groups within the phonemes.

Table 5.10 summarises the pairwise comparison in a matrix. The highlighted figures show the significances within the age groups.

Table 5.9: Pairwise comparison of the interaction of phoneme*age group in F1 (LU) (lsmeans with Tukey, significance level α = 0.05)

Interaction	Estimate	Std. Error	t value	Pr(>|t|)	
fleece,young - kit.young	-0.419	0.035	-12.112	<0.001	***
fleece,older - kit,young	-0.311	0.076	-4.086	0.002	**
fleece,middle - kit,young	-0.275	0.059	-4.625	<0.001	***
kit,older - kit,young	-0.193	0.073	-2.638	0.105	
fleece,middle - kit,middle	-0.139	0.036	-3.852	0.002	**
kit,middle - kit,young	-0.136	0.055	-2.479	0.148	
fleece,older - kit,older	-0.118	0.051	-2.342	0.179	
fleece,middle - kit,older	-0.082	0.079	-1.042	0.902	
fleece,middle - fleece,older	0.036	0.075	0.479	1.000	
kit,middle - kit,older	0.057	0.076	0.750	0.974	
fleece,older - fleece,young	0.108	0.072	1.483	0.676	
fleece,middle - fleece,young	0.144	0.055	2.624	0.109	
kit,middle - fleece,older	0.175	0.078	2.232	0.238	
kit,older - fleece,young	0.227	0.076	2.962	0.047	*
kit,middle - fleece,young	0.283	0.059	4.806	<0.001	***

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'

Table 5.10: Matrix of the pairwise comparison of F1 (LU) for the interaction of phoneme*age group for FLEECE and KIT (lsmeans with Tukey, significance level α = 0.05)

Interaction	FLEECE* young	FLEECE* middle	FLEECE* older	KIT* young	KIT* middle
FLEECE*middle	-				
FLEECE*older	-	-			
KIT*young	0.42	0.28	0.31		
KIT*middle	0.28	0.14	-	-	
KIT*older	0.23	-	-	-	-

Table 5.11 shows the variables that yielded statistically significant results for F2. These are the variables *phoneme, age group, speech style, preceding context* (coronals, non-coronals)

and the interactions of *phoneme*age group* and *phoneme*gender*. In terms of the variables with binary labels, KIT is produced 0.34 LU backer than FLEECE. Preceding non-coro-nals are produced 0.10 LU fronter than coronals, which is line with Thomas' (2011: 49) observation. *Gender, number of vowels in L1* and *age group*gender* showed no statistical significance. $R^2c = 0.35$ is again moderately high. The outcome is explained by 20% of the fixed factors and 15% of the random factors. The Tukey post-hoc test for *speech style* yielded significances only for the comparison WL-IN. The other two comparisons produced $p = 0.9$ (ReP–IN) and $p = 0.2$ (ReP–WL).

Table 5.11: F2 regression results for FLEECE (N = 766) and KIT (N = 731) (ph = phoneme, ag = age group, pc = preceding context)

```
Scaled residuals:
                      Min       1Q     Median     3Q       Max
                    -6.4200   -0.5414   0.0723   0.6211   2.9005
Random effects:
 Groups              Variance   Std.Dev.
 word label          0.0375     0.1937
 speaker name        0.0050     0.0709
 Residual            0.1896     0.4355
Fixed effects:
                     Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)           1.4190    0.0542   207.40   26.202    <2e-16 ***
 ph kit               -0.3356    0.0604   461.40   -5.556  4.67e-08 ***
 ag young              0.1575    0.0431    83.10    3.653    <0.001 ***
 pc non-coronal        0.1025    0.0401   280.80    2.556     0.011 *
 style WL              0.1324    0.0505   167.50    2.624     0.010 **
 ph kit:ag young      -0.3422    0.0513  1437.00   -6.669  3.68e-11 ***
 ph kit:gender male    0.1089    0.0474  1424.00    2.299     0.022 *

R²m (fixed effects) = 0.20
R²c (fixed + random effects) = 0.35

Post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 WL - IN == 0           0.1324     0.0505       2.624     0.022 *

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

Table 5.12 shows the statistically significant results of the interaction of *phoneme*age group* in a matrix. The figures highlighted in grey are the differences within the age groups. All age groups differ in their pronunciation of FLEECE and KIT albeit to different degrees. The young speakers show the greatest distinction between the two vowels with FLEECE 0.62 LU fronter than KIT. Young speakers produce KIT significantly backer than the other two groups, and they produce FLEECE significantly fronter that the middle group. Within the phonemes, there are no statistically significant differences between the middle and the older age group.

Table 5.12: Pairwise comparison of F2 (LU) for the interaction of phoneme*age group for FLEECE and KIT

Interaction	FLEECE* young	FLEECE* middle	FLEECE* older	KIT* young	KIT* middle
FLEECE*middle	-0.16				
FLEECE*older	-	-			
KIT*young	-0.62	-0.47	-0.47		
KIT*middle	-0.44	-0.28	-0.29	0.18	
KIT*older	-0.36	-0.21	-0.21	0.26	-

The statistically significant results of the interaction of *phoneme*gender* across the phonemes are shown in Table 5.13 within the genders. Of the six possible combinations, only the values for FLEECE*male and FLEECE*female did not differ (p = 0.99) (not given in the tables). Both phonemes are produced differently across the genders; male and female speakers pronounce FLEECE much more fronted than KIT and males produce KIT somewhat fronter than females.

Table 5.13: Pairwise comparison of F2 (LU) for the interaction of phoneme*gender for FLEECE and KIT

Interaction	FLEECE*female	FLEECE*male	KIT*female
FLEECE*male	-		
KIT*female	-0.43	-0.43	
KIT*male	-0.31	-0.32	0.11

Due to the computational constraints of the regression analysis used, the models were run separately for the two formants. A comprehensive view, of course, must include both dimensions – vowel height and vowel frontness. Considering both, it can be said that the linguistic factor *preceding context* has a minor impact on pronunciation. Of the social factors, *age group* has the most decisive effect, i.e. speakers of the young group showed the most distinct realisation of KIT and FLEECE.

5.2.1.2 Vowel duration

Figure 5.4 shows the normalised vowel duration of FLEECE and KIT represented by boxplots. The FLEECE vowel does not contain word-final tokens followed by a pause. The median within the box is higher in FLEECE than in KIT. The FLEECE box itself is bigger, and the upper quartile has a higher peak. It is evident that FLEECE is generally pronounced longer than KIT. Leaving the outliers aside, FLEECE values range roughly between -85 ms and 75 ms whereas the KIT values range between -85 ms and 40 ms. The end of the lower quartile is basically the same in both phonemes meaning that the tokens towards the lower end (i.e., tokens with a low vowel duration) were produced similarly short. Strikingly, there are many outliers in both phonemes thus hinting at a high degree of variation in vowel length. Table 5.14 reflects the above description. With a mean of -9.52 ms, FLEECE is on average 21 ms longer than KIT. Their SDs are similarly high.

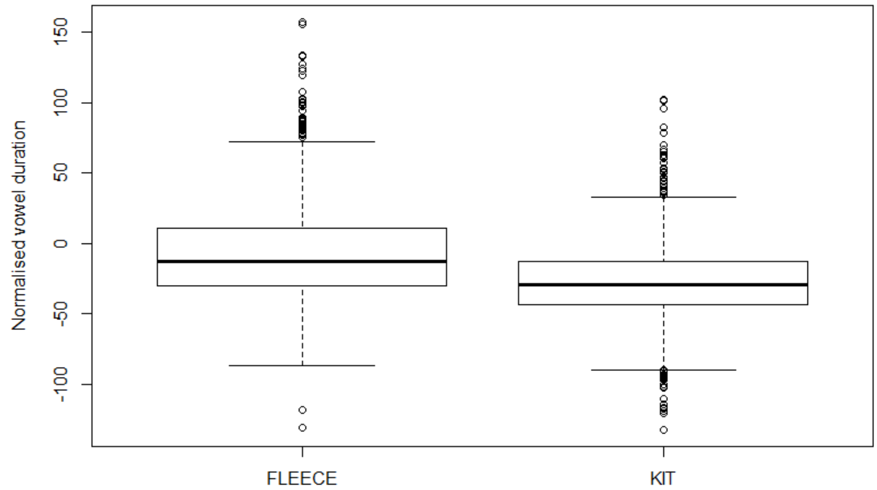

Figure 5.4: Normalised vowel duration (ms) of the high front cluster

Table 5.14: Token numbers, normalised vowel durations and SDs of the high front cluster

Phoneme	N	Mean (ms)	SD
FLEECE	752	-9.52	35.53
KIT	731	-30.42	30.97

The fixed factors entered into the model were *phoneme, speech style, gender, age group, following voicing* (voiced, voiceless, pause/gap), following *manner* (fricative, affricate, plosive, pause/gap) and the interactions of *age group*gender, phoneme*gender, phoneme*age group* and *phoneme*speech style*. The variables that yielded statistically significant results are shown in Table 5.15. These are *phoneme, gender, age group, following manner* and the interactions of *age group*gender* and *phoneme*speech style.* Interestingly, the variable *following voicing* was not statistically significant ($p < 0.4$), and neither was the interaction of *phoneme*age group* ($p = 0.06$). $R^2c = 0.42$ implies a moderate fit. The fixed factors (23%) and the random factors

(19%) can account for 42% of the variability of vowel duration. The post-hoc test for the variable *following manner* showed that vowels followed by fricatives were produced longer than those followed by affricates and plosives. This result is confirmed by Peterson & Lehiste (1960: 702).

Table 5.15: Regression results of duration for FLEECE (N = 752) and KIT (N = 731) (ph = phoneme, ag = age group)

```
Scaled residuals:
                      Min        1Q      Median      3Q        Max
                   -3.3912   -0.5786   -0.1108   0.4676   4.4415
Random effects:
 Groups             Variance   Std.Dev.
 word label         248.70     15.77
 speaker name         0.00      0.00
 Residual           730.00     27.02

Fixed effects:
                    Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)         -15.010    6.327    168.50   -2.373     0.019 *
 gender male          -5.022    2.380   1425.90   -2.111     0.035 *
 ph kit              -18.179    3.914    357.10   -4.644  4.81e-06 ***
 foll manner fricat   21.194    6.540    134.40    3.241     0.002 **
 ag young:ph kit      -6.193    3.186   1426.10   -1.944     0.052
 ph kit:style WL     -18.231    7.240    229.60   -2.518     0.012 *
 ag young:gend male    9.511    3.228   1426.40    2.946     0.003 **

R²m (fixed effects) = 0.23
R²c (fixed + random effects) = 0.42

Post-hoc Tukey, confidence level 0.95:
                     Estimate   Std. Error   z value   Pr(>|z|)
 fric - affric == 0   21.194      6.540        3.241     0.005 **
 plosive - fric == 0  -8.917      3.290       -2.710     0.027 *

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

The post-hoc Tukey test for the interaction of *phoneme*age group* in Table 5.16 shows that there are significant differences between FLEECE and KIT across and within the age groups in that FLEECE is generally produced longer than KIT. The age groups do not differ within the phonemes, i.e. neither do they differ in the length of FLEECE nor of KIT. The figures highlighted in grey are the differences within the age groups.

Table 5.16: Pairwise comparison of duration (ms) for the interaction of phoneme*age group for FLEECE and KIT

Interaction	FLEECE* young	FLEECE* middle	FLEECE* older	KIT* young	KIT* middle
FLEECE*middle	-				
FLEECE*older	-	-			
KIT*young	-28.39	-23.99	-19.80		
KIT*middle	-26.59	-22.20	-18.00	-	
KIT*older	-31.95	-27.55	-23.36	-	-

The post-hoc test for the interaction of *age group*gender* yielded only two statistically significant results: Young males produced both phonemes on average 11.88 ms ($p = 0.02$) longer than older males, and females of the middle group produced both phonemes 10.85 ms ($p = 0.05$) longer than older males. It can thus be assumed that age in combination with gender has little influence on vowel duration. The post-hoc test for the interaction of *phoneme*speech style* in Table 5.17 shows that in WL style, all FLEECE tokens were produced significantly longer (between 39 ms and 47 ms) than the KIT tokens. Within KIT, ReP and IN tokens were longer than those of WL. It seems that the speakers consciously pronounced KIT vowels in citation style as lax vowels. In ReP and IN style, they perhaps paid less attention. By contrast, there were no differences in speech style within FLEECE.

Table 5.17: Pairwise comparison of duration (ms) for the interaction of phoneme*speech style for FLEECE and KIT

Interaction	FLEECE*IN	FLEECE*ReP	FLEECE*WL	KIT*IN	KIT*ReP
FLEECE*ReP	-				
FLEECE*WL	-	-			
KIT*IN	-20.63	-	-		
KIT*ReP	-17.87	-	-	-	
KIT*WL	-47.02	-43.60	-38.86	-26.39	-29.15

5.2.1.3 KIT split

It was investigated whether the KIT split, a phenomenon of White South African English (see Section 3.2.2), also occurred in BSAE. Van Rooy & Van Huyssteen (2000) divided the KIT vowel into KIT and SIT in their investigation of Tswana-English. Fourteen percent of the SIT tokens were realised as [ɛ], indicating that the KIT split is not an exclusively WSAE feature.

Since the KIT split involves a horizontal rather than a vertical shift, only F2 was modelled. Of the regular parameters, *preceding context* was replaced by *adjacent context* (high front vs. centralised). The interactions of *phoneme*gender* and *phoneme*age group* were replaced by *adjacent context*gender* and *adjacent context*age group*. The high front variant included all tokens in which KIT occurred

- in word-initial position (*it*),
- after /h/ (*hit*),
- adjacent to velar consonants (*kids, give, sick, thing*) and
- before post-alveolar /ʃ/ (*fish*).

The centralised variant comprised all other environments. The phonological contexts /l, n, r, w/ were omitted, however. The results are shown in Table 5.18. Adjacent context, age group and the interactions of *adjacent context*gender* and *adjacent context*age group* had significant influence on the split. KIT tokens of the high front variant are produced 0.32 LU fronter than those in the centralised variant.

The variables *gender, speech style, number of vowels in L1* and the interaction of *age group*gender* showed no statistical significance. Whereas WL style showed a significant difference towards IN in the initial model output, the Tukey test produced p-values between 0.09 and 0.9 meaning that speech style is not statistically significant. $R^2c = 0.50$ implies a moderate fit. 22% of the fixed factors and 28% of the random effects account for the variability in the KIT split.

Table 5.18: F2 regression results for KIT (N = 731) (ac = adjacent context, ag = age group, hf = high front)

```
Scaled residuals:
                       Min       1Q     Median     3Q       Max
                    -4.5758   -0.5581   0.0198   0.5787   2.6348
Random effects:
 Groups              Variance   Std.Dev.
 word label          0.0434     0.2084
 speaker name        0.0525     0.2291
 Residual            0.1710     0.4136
```

```
Fixed effects:
                      Estimate  Std. Error   df     t value   Pr(>|t|)
(Intercept)            1.0195    0.0862     63.20   11.823    <2e-16 ***
 ac high front         0.3156    0.0782    296.60    4.031   7.07e-05 ***
 ag young             -0.3963    0.0876     48.10   -4.521   4.02e-05 ***
 style WL              0.1362    0.0656    127.60    2.075      0.040 *
 ac hf:ag young        0.3654    0.0693    764.40    5.276   1.72e-07 ***
 ac hf:gender male    -0.1778    0.0634    750.60   -2.803      0.005 **

R²m (fixed effects) = 0.22
R²c (fixed + random effects) = 0.50

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

The interaction of *adjacent context*age group* is shown in Table 5.19. All age groups differ in their pronunciation of words in the two contexts: If KIT occurred word-initially, after /h/, before /ʃ/ and adjacent to velar consonants, the speakers pronounced a high front vowel. In all other contexts, they produced a centralised variant. With 0.59 LU, young speakers have the greatest distance between the variants. In terms of the centralised pronunciation, the young group is 0.4 LU backer than the other age groups. The distance between the middle and the older group is zero with $p = 1$. It is interesting that there were no significances between the age groups in the high front variant. Here, the values ranged between 0.03 and 0.15 LU with $p > 0.80$.

Table 5.19: Pairwise comparison of F2 (LU) for the interaction of adjacent context*age group for KIT

Interaction	central.* young	central.* middle	central.* older	hi. front* young	hi. front* middle
centralised*middle	0.40				
centralised*older	0.40	-			
high front*young	0.59	-	-		
high front*middle	0.62	0.23	-	-	
high front*older	0.75	-	0.34	-	-

Figure 5.5 visualises the above results including 1 SD. On average, centralised KIT and the high front variant are distinct. The latter is produced very similarly by all age groups. The SDs range between 0.41 and 0.44 and have a very similar spread. Centralised KIT in the middle and older group have practically the same means and SDs and differs considerably from the

young participants. The young group has an SD of 0.64, which suggests great variation among the young speakers.

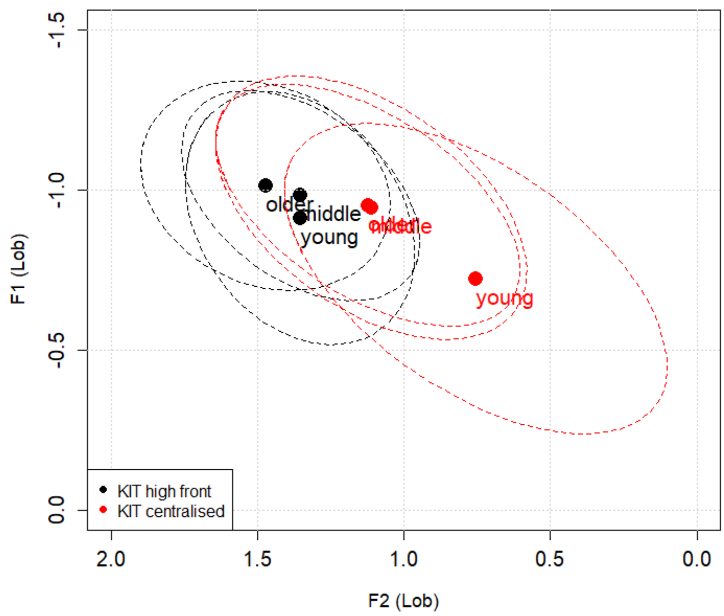

Figure 5.5: **KIT high front (N = 316) and KIT centralised (N = 499) by age group with 1 SD**

In order to check the distance between high front KIT and centralised KIT in each individual speaker, the Euclidian distance (ED) between the means of the formants F1 and F2 for the two variants were calculated. Although the Euclidian distance can neither provide information about the degree of overlap nor indicate statistical significance, it is a useful measure for determining the relative placement of two vowel points in a two-dimensional vowel space (Fabricius 2007: 303; Nycz & Hall-Lew 2014: 3). The following formula was used to calculate the Euclidian distance (adapted from Fabricius 2007: 302):

$$ED = \sqrt{(F1_{high\ front} - F1_{centralised})^2 + (F2_{high\ front} - F2_{centralised})^2}$$

The smaller the ED value, the closer the two KIT variants are. The values ranged between 0.03 LU and 1.45 LU. Twenty speakers had a distance of 0.49 LU and higher, hinting at a long distance between both positions. In nine of them (eight young speakers and an older speaker), the sum of the SD values for F2 high front and F2 centralised was smaller than their Euclidian distance, which indicates that these speakers produced two distinct KIT variants. The interaction of adjacent *context*gender* is shown in Table 5.20. Also, within the genders, the two variants were pronounced differently.

Table 5.20: Pairwise comparison of F2 (LU) for the interaction of adjacent context*gender for KIT

Interaction	centralised*female	centralised*male	high front*female
centralised*male	-		
high front*female	0.48	0.32	
high front*male	0.46	0.30	-

The KIT split, so far reported for WSAE has made its way to BSAE. Speakers of all age groups and both sexes show this phenomenon. The fact that the KIT split has occurred also in the older age group suggests that this feature must have been around for a longer time, considering that language behaviour is set up during childhood and adolescence and cannot be easily changed in adulthood. Two outcomes are particular interesting: First, the young speakers were most evidently distinct in the pronunciation of the centralised variant in two respects – compared to the high front variant of the same age group and compared to the centralised variant of the other two age groups. Second, female speakers show a larger distance between the two variants than males. The reason why the KIT split has been assigned only to WSAE so far may be due to the fact that it has not been investigated in detail in other SAE varieties. Van Rooy & Van Huyssteen (2000) found incidents of a centralised variant in the English of young Tswana speakers, but they did not refer to this finding as KIT split.

5.2.1.4 Summary of the key findings

Regarding vowel quality, the variables *phoneme*, *gender*, *age group, speech style* and *preceding context* and the interaction of *phoneme*age group* were decisive. KIT is generally produced lower and backer than FLEECE hinting at two distinct vowels, which applies to both genders and the age groups. The young group shows the longest distance between KIT and FLEECE. Speech style influenced vowel height, but only in the comparison WL-IN. With regard to the KIT split, speakers of all age groups and both genders perform it although to different degrees. Concerning the preceding context, non-coronals are produced fronter and slightly higher than coronals. For the vowel quality of FLEECE, the phoneme /i/ is proposed for all age groups. The vowel quality of KIT may be /ɪ/ for the young group and /i/ for speakers of the middle and older group.

Regarding vowel duration, the variables *phoneme, gender, age group* and *following manner* and the interactions of *age group*gender* and *phoneme*speech style* were statistically significant. Interestingly, the variable *following voicing* did not play a role. FLEECE is generally produced longer than KIT, which is true for all genders and all age groups. Age group in combination with gender is statistically significant, but only in very few combinations. It therefore seems to have only little influence on vowel duration. In terms of consonant manner, vowels followed by fricatives were produced longer than those followed by affricates and plosives. Speech style has some influence on vowel duration. In WL, FLEECE is produced longer than KIT and KIT is produced longest in WL compared to the other styles.

5.2.2 The mid front vowels

5.2.2.1 Vowel quality

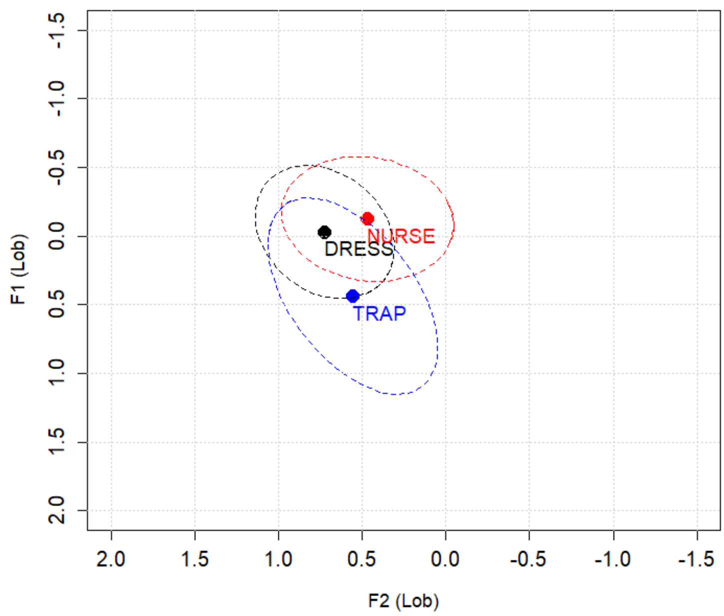

Figure 5.6: Mid front cluster across all variables with 1 SD

The means of DRESS, NURSE and TRAP are close together and their SD overlap is large, which is particularly obvious with DRESS and NURSE (see Figure 5.6 and Table 5.21). NURSE lies within 1 SD of DRESS, and the TRAP mean scratches at the SD line of DRESS. NURSE has a great F2 spread indicating also central realisations. Of this cluster, TRAP has the largest F1 expansion, covering large parts of the DRESS area [e] but also its prescriptively inherent place as a fairly-open unrounded front vowel [æ]. Apart from that, some speakers produce it also lowered and retracted.

As illustrated in Table 5.21, the vowel overlap is very high with

- BC = 0.93, Pillai = 0.09, $p < 0.001$ for DRESS-NURSE
- BC = 0.89, Pillai = 0.13, $p < 0.001$ for DRESS-TRAP and

- BC = 0.81, Pillai = 0.22, p < 0.001 for TRAP-NURSE

and suggests that these vowels are not completely distinct. As can also be seen in Figure 5.7, a large overlap was between DRESS and NURSE. Again, some young speakers show very little overlap. These were TYF04 (0.07), TYM07 (0.12), ZYM06 (0.17), SYF10 (0.17) and XYF03 (0.26). In the distribution of DRESS and TRAP, seven young speakers (TYF04, SYF05, XYF03, ZYM04, ZYM09, TYM07, SYF10) showed the lowest BCs, ranging between 0.00 and 0.27 (the next value was 0.53).

Table 5.21: Token numbers, formant means and SDs of the mid front cluster

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
DRESS	1848	-0.03	0.48	0.72	0.42
TRAP	1570	0.44	0.72	0.56	0.51
NURSE	1086	-0.13	0.45	0.47	0.51

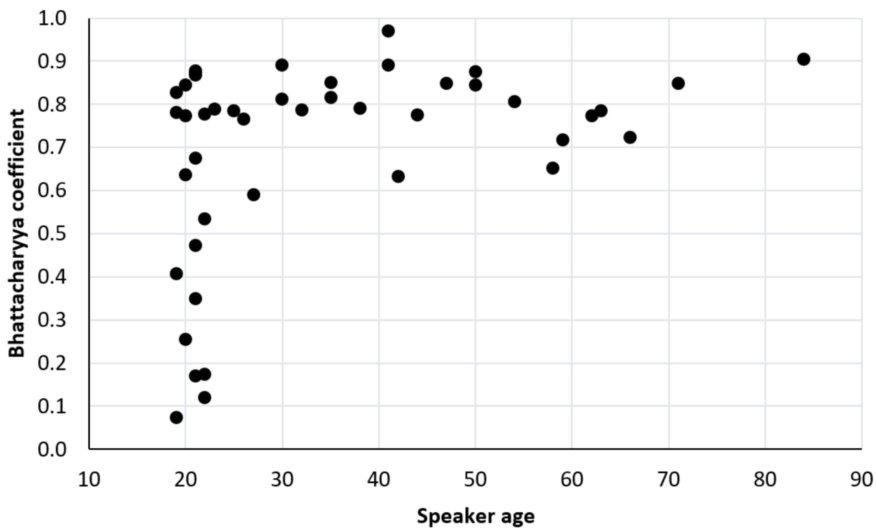

Figure 5.7: Bhattacharyya coefficients for DRESS (N = 1190) and NURSE (N = 815) in IN style separated by speaker age

A regression analysis was carried out for F1 and F2. The fixed factors for both runs were *phoneme* (DRESS, TRAP, NURSE), *gender, age group, speech style, number of vowels in L1, preceding context* (coronals, non-coronals) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. The variables that yielded statistically significant results for F1 are shown in Table 5.22. These are the variables *phoneme, age group, speech style* and the interaction of *phoneme*age group*. R^2 values are moderately high with 0.29 and 0.50. The outcome can be explained by 29% of the fixed factors and 21% of the random effects. The post-hoc test showed that speech style influenced vowel height; tokens in WL style were pronounced lower than those in IN and ReP style.

Table 5.22: F1 regression results for DRESS (N = 1015), NURSE (N = 692) and TRAP (N = 806) (ph = phoneme, ag = age group)

```
Scaled residuals:
                      Min       1Q      Median     3Q       Max
                   -3.7198   -0.6193   -0.0444   0.5574   5.0271
Random effects:
 Groups             Variance   Std.Dev.
 word label         0.0809     0.2844
 speaker name       0.0069     0.0831
 Residual           0.2072     0.4552

Fixed effects:
                    Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)         -0.1435    0.0422   235.60   -3.403     0.001 ***
 ph trap              0.5093    0.0542   779.70    9.390    <2e-16 ***
 ag young            -0.1184    0.0429    90.30   -2.763     0.007 **
 style WL             0.2868    0.0576   716.20    4.974   8.2e-07 ***
 ph trap:ag young     0.3758    0.0495  2398.00    7.595   4.4e-14 ***

R²m (fixed effects) = 0.29
R²c (fixed + random effects) = 0.50

Post-hoc Tukey, confidence level 0.95:
                     Estimate   Std. Error   z value   Pr(>|z|)
 WL - IN == 0          0.2868     0.0576       4.974    <0.001 ***
 WL - ReP == 0         0.2740     0.0716       3.826    <0.001 ***

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

The interaction of *phoneme*age group* is given in Table 5.23. The figures highlighted in grey are the differences within the age groups. Of the 36 possible combinations, 21 were statistically significant. What becomes obvious is that the significances between the phonemes only occur between TRAP and DRESS, TRAP and NURSE but not between DRESS and NURSE. Generally, speakers of all age groups produced TRAP lower than DRESS and NURSE. Young speakers showed the largest distance between TRAP and DRESS as well as TRAP and NURSE. The vowel height of DRESS and NURSE is very similar within the age groups and across them. Regarding DRESS, there is a significant difference in the vowel height between young and older.

Table 5.23: Pairwise comparison of F1 (LU) for the interaction of phoneme*age group for DRESS, NURSE and TRAP

Interaction	DRESS* young	DRESS* middle	DRESS* older	NURSE* young	NURSE* middle	NURSE* older	TRAP* young	TRAP* middle
DRESS*mid.	-							
DRESS*old.	0.18	-						
NURSE*you.	-	-	-					
NURSE*mid.	-	-	-	-				
NURSE*old.	-	-	-	-	-			
TRAP*you.	0.88	0.76	0.70	0.86	0.78	0.83		
TRAP*mid.	0.63	0.51	0.45	0.61	0.52	0.57	-0.26	
TRAP*older	0.65	0.53	0.47	0.63	0.55	0.59	-0.23	-

The model for F2 yielded significances only for *phoneme, age group* and the interaction of both (Table 5.24). R^2c is moderately high, but the value for the mixed effects is very low.

Table 5.24: F2 regression results for DRESS (N = 1015), NURSE (N = 692) and TRAP (N = 806) (ph = phoneme, ag = age group)

```
Scaled residuals:
                      Min       1Q     Median     3Q       Max
                   -4.0976   -0.5591   0.0279   0.5958   4.1706
Random effects:
 Groups             Variance   Std.Dev.
 word label         0.0705     0.2656
 speaker name       0.0061     0.0780
 Residual           0.1334     0.3652
Fixed effects:
                    Estimate  Std. Error   df     t value    Pr(>|t|)
 (Intercept)          0.6652    0.0396   289.80   16.787     <2e-16 ***
 ph nurse            -0.0997    0.0495   573.30   -2.015      0.048 *
 ph trap             -0.1552    0.0466   768.80   -3.331      0.001 ***
 ag young             0.1085    0.0370    85.30    2.931      0.004 **
 ph nurse:ag young   -0.4330    0.0419  2396.70  -10.340     <2e-16 ***
 ph trap:ag young    -0.1135    0.0400  2370.00   -2.841      0.004 **

R²m (fixed effects) = 0.10
R²c (fixed + random effects) = 0.43

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

The matrix in Table 5.25 shows the pairwise comparison of the interaction of *phoneme*age group* for F2. The figures highlighted in grey are the differences within the age groups. The outcome for vowel frontness is more diffuse than those of vowel height, but four results are evident: First, within the age groups, young speakers show the biggest horizontal distance between the phonemes (DRESS-NURSE 0.50 LU, DRESS-TRAP 0.26 LU and TRAP-NURSE 0.27 LU). Second, the difference in the pronunciation of NURSE in the young group is statistically significant compared to the realisation of DRESS, TRAP and NURSE in the middle and older age group; young speakers produce NURSE significantly backer than any other age group or combination (see also Figure 5.8). Third, it seems that DRESS is produced fronter than NURSE and TRAP by all age groups although it is only statistically significant in the young group. Fourth, the comparison of DRESS as well as of TRAP across the age groups did not yield statistically significant differences.

Table 5.25: Pairwise comparison of F2 (LU) for the interaction of phoneme*age group for DRESS, NURSE and TRAP

Interaction	DRESS* young	DRESS* middle	DRESS* older	NURSE* young	NURSE* middle	NURSE* older	TRAP* young	TRAP* middle
DRESS*mid.	-							
DRESS*old.	-	-						
NURSE*you.	-0.50	-0.42	-0.48					
NURSE*mid.	-0.21	-	-	0.32				
NURSE*old.	-0.23	-	-	0.30	-			
TRAP*you.	-0.26	-	-0.21	0.27	-	-		
TRAP*mid.	-0.26	-0.15	-0.21	0.27	-	-	-	
TRAP*older	-0.22	-	-	0.31	-	-	-	-

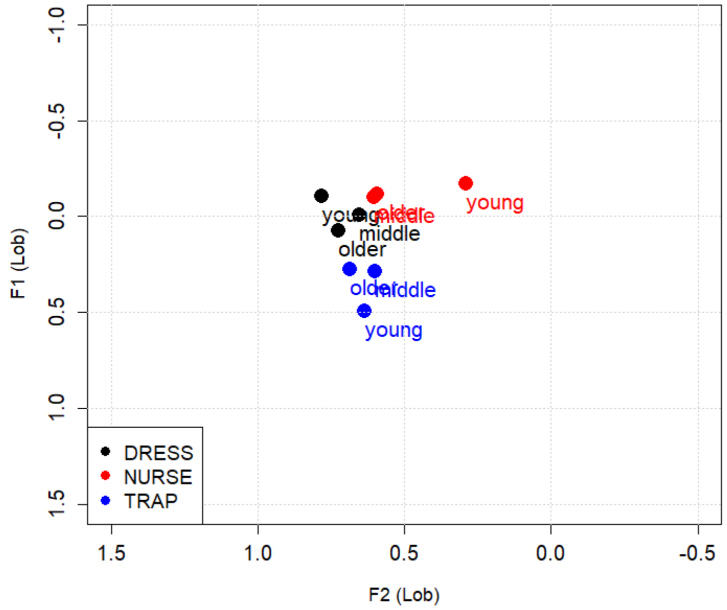

Figure 5.8: Mid front cluster by age group

5.2.2.2 Vowel duration

Figure 5.9 shows the normalised vowel duration of the mid front cluster. The NURSE vowel does not contain word-final tokens followed by a pause. The long vowel in NURSE is produced longer than those in the two short vowels. This is marked by the upper quartile, which ends slightly below 100. Two features are prominent: First, the similar end of the lower quartile and second, the numerous outliers, especially in the upper part. Moreover, there is a high degree of variation in vowel length in this cluster.

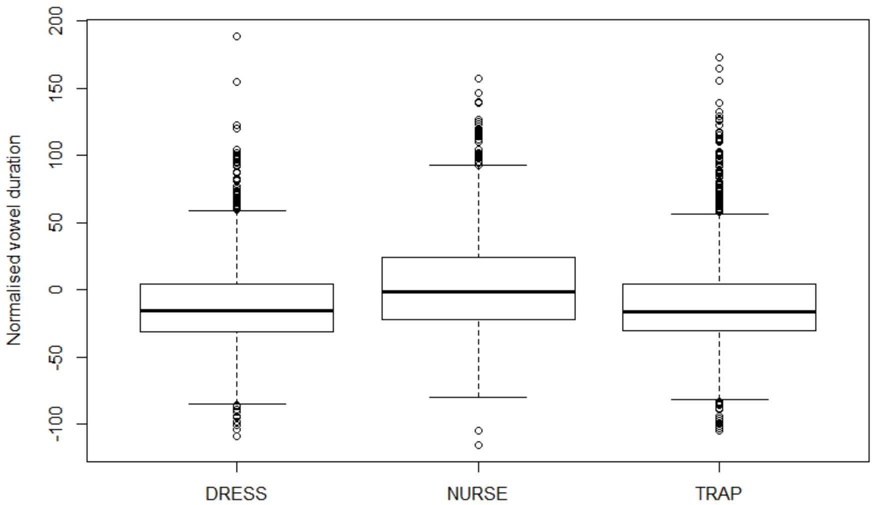

Figure 5.9: Normalised vowel duration (ms) of the mid front cluster

Table 5.26 displays the means and SDs. The normalised mean of NURSE is with 5.37 ms longer than those of DRESS and TRAP, which both yielded negative values. Again, the SDs of the three phonemes are equally high and also high in absolute figures.

Table 5.26: Token numbers, normalised vowel durations and SDs of the mid front cluster

Phoneme	N	Mean (ms)	SD
DRESS	1015	-12.96	32.14
NURSE	686	5.37	37.37
TRAP	806	-7.25	39.13

The fixed factors entering the model were *phoneme, speech style, gender, age group, following voicing* (voiced, voiceless, pause/gap), *following manner* (fricative, affricate, plosive, pause/gap) and the interactions of *age group*gender, phoneme*gender, phoneme*age group* and *phoneme*speech style*. The variables that yielded statistically significant results are shown in Table 5.27. These are the variables *phoneme* and *age group* and the interactions of *phoneme*age group* and *phoneme*speech style*. The variables *following voicing* and *following manner* were not statistically significant. $R^2c = 0.44$ implies a moderate fit, but the fixed factors account for only 8% for the outcome.

Table 5.27: Regression results of duration for DRESS (N = 1015), NURSE (N = 686) and TRAP (N = 806) (ph = phoneme, ag = age group)

```
Scaled residuals:
                    Min       1Q      Median     3Q       Max
                  -3.5313   -0.5796   -0.1219   0.4554   7.0481
Random effects:
 Groups            Variance   Std.Dev.
 word label        437.95     20.930
 speaker name       18.93      4.350
 Residual          716.56     26.770

Fixed effects:
                     Estimate Std. Error   df     t value   Pr(>|t|)
 (Intercept)          -11.246   2.719     378.30   -4.137   4.35e-05 ***
 ph nurse              18.309   3.855     671.10    4.750   2.49e-06 ***
 ph trap               10.560   3.710     760.40    2.846      0.004 **
 ag young              -6.745   2.427     110.20   -2.779      0.006 **
 ph nurse:ag young      7.564   3.089    2392.70    2.449      0.014 *
 ph nurse:style ReP   -19.920   7.402    2448.10   -2.691      0.007 **
 ph trap:style WL     -14.542   7.186    1511.20   -2.024      0.043 *

R²m (fixed effects) = 0.08
R²c (fixed + random effects) = 0.44

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

The pairwise comparison of the interaction of *phoneme*age group* is shown in Table 5.28. In seven cases, DRESS is pronounced significantly shorter than NURSE. Although only statistically significant results are displayed, it should be mentioned that NURSE was always pronounced longer than DRESS and TRAP across the age groups. The differences ranged between 5.43 and 20.54 ms. The young group showed significant differences between all three phonemes, and the middle group only between DRESS and NURSE. The older group did not show any differences in this respect.

Table 5.28: Pairwise comparison of duration (ms) for the interaction of phoneme*age group for DRESS, NURSE and TRAP

Interaction	DRESS* young	DRESS* middle	DRESS* older	NURSE* young	NURSE* middle	NURSE* older	TRAP* young	TRAP* middle
DRESS*mid.	-							
DRESS*old.	-	-						
NURSE*you.	24.84	18.02	15.42					
NURSE*mid.	23.90	17.07	14.48	-				
NURSE*old.	20.62	-	-	-	-			
TRAP*you.	11.62	-	-	-13.22	-	-		
TRAP*mid.	15.28	-	-	-	-	-	-	
TRAP*older	-	-	-	-	-	-	-	-

The pairwise comparison of the interaction of *phoneme*speech style* (Table 5.29) shows seven significant results but does not display a pattern. Within the styles, only DRESS and NURSE in IN style and NURSE and TRAP in WL style were significantly different in that DRESS is produced shorter than NURSE in IN, and TRAP is produced shorter than NURSE in WL.

Table 5.29: Pairwise comparison of duration (ms) for the interaction of phoneme*speech style for DRESS, NURSE and TRAP

Interaction	DRESS* IN	DRESS* ReP	DRESS* WL	NURSE* IN	NURSE* ReP	NURSE* WL	TRAP* IN	TRAP* ReP
DRESS*ReP	-							
DRESS*WL	-	-						
NURSE*IN	18.79	17.32	25.83					
NURSE*ReP	-	-	-	-				
NURSE*WL	-	-	-	-	-			
TRAP*IN	-	-	-	-	-	-		
TRAP*ReP	16.80	-	-	-	-	-	-	
TRAP*WL	-	-	-	-30.53	-	-27.41	-21.58	-

5.2.2.3 NURSE spelling

The various graphemic representations of the NURSE vowel (e.g. *word, learn, first, turn, nerd, journey*) suggest an influence on the pronunciation of these lexical subsets. Simo Bobda (2000b) provides a detailed auditory account of the pronunciation of NURSE in selected countries of sub-Saharan Africa. He found a very high degree of variability and considers NURSE "the most distinguishing parameter in the regional, national and even ethnic identification of a speaker" (Simo Bobda 2000b: 41). He provides examples of the three major Nigerian English varieties Hausa English, Yoruba English and Igbo English in which the NURSE paradigm splits into several subsets due to spelling and position in the word. These findings are supported by Schmied (2004: 925), who found examples of spelling pronunciation in East African English in which the NURSE vowel assumes "the sound value "suggested" by the orthographic symbol that represents it (e.g. [adʒ] for *urge* vs. [heːd] for *heard*)." For Southern Africa – comprising the countries Zimbabwe, Botswana, Namibia, South Africa, Lesotho and Swaziland – Simo Bobda (2000b) attests only [ɛ] for all spellings.

To test the previous results, a model for spelling was run for F1 and F2. The fixed factors for both runs was *spelling, gender, age group,* and the interaction of *spelling*age group* and *spelling*gender.* The variable *speech style* was excluded because not every spelling variant occurred in each speech style. Of the 781 NURSE tokens, 775 entered the model; the variant

<our> occurred only six times in the dataset and was thus left out because of the small token number. The NURSE tokens split as follows:

Table 5.30: Spelling variants of NURSE (N = 781)

NURSE spelling	N	Mean F1	SD F1	Mean F2	SD F2
ear	58	-0.03	0.45	0.60	0.44
er	252	-0.22	0.44	0.51	0.50
ir	189	-0.10	0.39	0.44	0.41
our	6	0.08	0.85	0.72	0.49
ur	193	-0.02	0.47	0.44	0.52
wor	83	-0.26	0.45	0.36	0.65

The results for F1 yielded no statistical significance at all, and the results for F2 (see Table 5.31) only for the comparison of *ear*- and *wor*-spelling with an estimate of 0.28 LU ($p = 0.03$). The other p-values ranged between 0.3 and 1.0. The *ear*-spelling is the most fronted and *wor*-spelling the backest realisation. The estimates of the pairwise comparison between <ir>, <er >and <ur> ranged between 0.003 and 0.017 and thus showed practically no differences. In terms of age group, the young participants realised NURSE significantly backer than the other groups as already described above. $R^2c = 0.44$ is moderately high, but $R^2m = 0.10$ relatively low. The interactions were not statistically significant.

Table 5.31: F2 regression results for NURSE spelling (N = 775) (sp = spelling, ag = age group)

```
Scaled residuals:
                   Min       1Q     Median     3Q       Max
                  -3.7119   -0.5703   0.0218   0.5707   3.4236
Random effects:
 Groups            Variance   Std.Dev.
 word label        0.0251     0.1585
 speaker name      0.0580     0.2408
 Residual          0.1432     0.3784
Fixed effects:
                   Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)         0.7192    0.0877   103.40    8.201   6.91e-13 ***
 sp wor             -0.2752    0.0956    88.15   -2.878      0.005 **
 ag young           -0.3088    0.0841    40.02   -3.669      0.001 ***

R²m (fixed effects) = 0.11
R²c (fixed + random effects) = 0.44
```

```
post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 wor - ear == 0        -0.2752    0.0956       -2.878     0.031 *

 young - middle == 0   -0.3088    0.0842       -3.669    <0.001 ***
 young - older == 0    -0.3628    0.1122       -3.233     0.004 **

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

Figure 5.10 zooms into the spelling variants of NURSE. DRESS and TRAP were added as reference points for the positions of the variants. Although the variants were produced very similarly, they nevertheless slightly differ in their means. The variant <ear> is closest to DRESS and may suggest spelling pronunciation as described for East Africa. A higher token number is needed to investigate whether this is the case or not, however. Other spelling pronunciations can be excluded. For example, <wor> is pronounced as a mid-front vowel and is thus far away from /ɔ/ as reported for Nigerian English varieties. The analysis showed that spelling had practically no influence on the pronunciation of NURSE in BSAE, which is in line with former findings for Southern Africa.

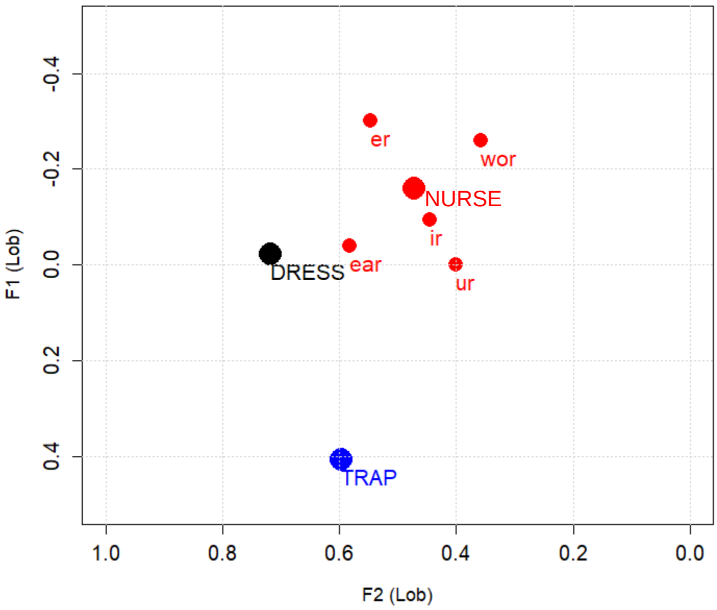

Figure 5.10: DRESS, TRAP and spelling variants of NURSE

5.2.2.4 Summary of the key findings

Regarding vowel quality, the variables *phoneme, age group, speech* and *style* and the interaction of *phoneme*age group* were decisive. The differences in vowel height could be found between TRAP and DRESS, and TRAP and NURSE, but not between DRESS and NURSE. In this respect, the young speakers showed the biggest distance between TRAP and the other two phonemes.

Vowel height of DRESS and NURSE was very similar in all age groups. Tokens in WL style were pronounced lower than those in IN and ReP style. Regarding vowel frontness, only young speakers show statistically significant differences between all phonemes. DRESS is produced fronter than NURSE and TRAP by all age groups although the results were not always statistically significant. Young speakers produce the backest NURSE vowel. The comparison of DRESS across the age groups did not yield statistically significant differences. The same applies to TRAP. Since young speakers differed in their performance from speakers of the middle and older age group, different vowel qualities are suggested. For the young group, DRESS is a close-mid vowel /e/, NURSE is a central vowel but raised to /ɘ/ and TRAP is /æ/. For the middle and older group, the open-mid vowel /ɛ/ is assumed for all three phonemes.

Regarding vowel duration, the variables *phoneme* and *age group* and the interactions of *phoneme*age group* and *phoneme*speech style* were statistically significant. The variables *following voicing* and *following manner* did not play a role. NURSE was always pronounced longer than DRESS and TRAP across the age groups, the results were not entirely statistically significant though. Speech style was decisive, but a pattern could not be discerned. The spelling variants of NURSE had practically no influence on the pronunciation, which corresponds with previous findings for Southern Africa.

5.2.3 The low central vowels

5.2.3.1 Vowel quality

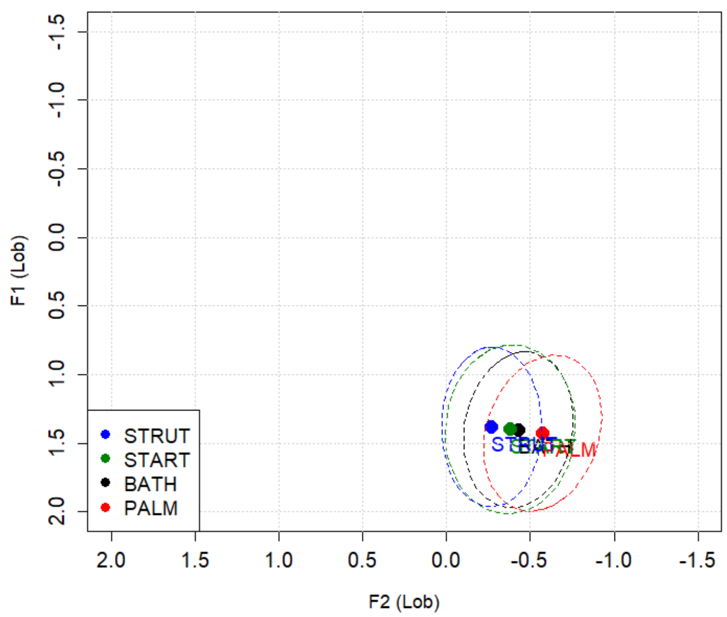

Figure 5.11: Low central cluster across all variables with 1 SD

Figure 5.11 shows the low central vowel cluster. The SDs of STRUT, BATH and START overlap very strongly. STRUT is slightly fronter than the other two phonemes. BATH and START can be regarded as one phoneme, a centralised and slightly raised open back unrounded [ɑ̈]. The visual evidence is backed by the figures in Table 5.32. The means and SDs of F1 are very close together and show identical vowel heights and SD spreads. The F2 values are less homogeneous but still similar. What is also obvious from the graph and the table is that PALM is produced at the same vowel height as the other vowels of this cluster, but backer than them. The overlap of the standard deviations is very big though.

Table 5.32: Token numbers, formant means and SDs of the low central cluster

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
BATH	519	1.40	0.57	-0.43	0.33
PALM	98	1.42	0.56	-0.56	0.36
START	596	1.40	0.61	-0.38	0.39
STRUT	1520	1.38	0.58	-0.27	0.30

The lexical set PALM consists of words with post-vocalic <lm> (i.e. phoneme /m/) but also of words with other post-vocalic contexts (e.g. /ð/ as in *father*). Bilabial consonants can have a lowering effect on both F1 and F2 (Thomas 2011: 101), which may be amplified when the vowel is preceded and followed by a bilabial (/p/ and /m/) and when the preceding bilabial is a nasal as is the case with the word *palm*. Another reason for the backing of /ɑ/ may lie in the fact that historically, postvocalic /l/ in words like *palm* and *calm* was actually pronounced (Mesthrie et al. 2015: 2). In order to check whether these consonant features possibly influence the pronunciation of PALM, a contour plot was created to compare the distribution of post-vocalic /m/ and other post-vocalic contexts as seen in Figure 5.12. The contour plot shows a visible overlap of the two subsets, but what is also obvious is that the /m/ subset is generally produced lower and backer than the second subset (see also Table 5.33).

Due to this bimodal distribution, PALM cannot be considered as a homogeneous set but rather as two subsets. The /m/ subset behaved exactly as described in the literature. The statistical model comparing both subsets of PALM, however, did not yield statistically significant differences. The F1 estimate was 0.21 LU with $p = 0.55$, and the F2 estimate was 0.19 LU with $p = 0.36$.

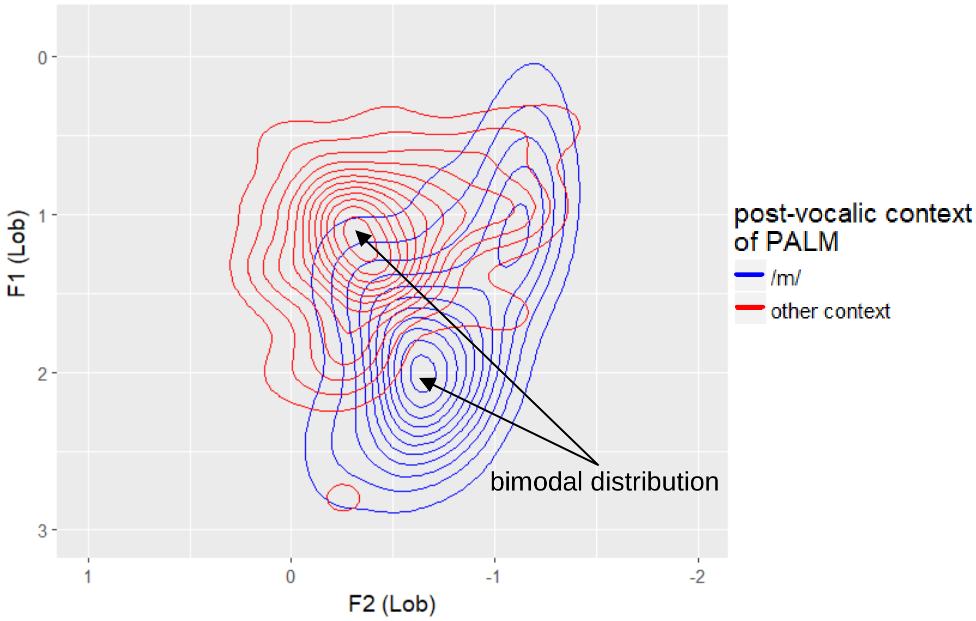

Figure 5.12: Contour plot of PALM subsets across all speakers and speech styles

Table 5.33: Token numbers, formant means and SDs of the PALM subsets

PALM	N	Mean F1	SD F1	Mean F2	SD F2
/m/	41	1.70	0.52	-0.74	0.28
other context	57	1.22	0.50	-0.43	0.35
Total	98	1.42	0.56	-0.56	0.36

The /m/ subset of PALM was excluded from further investigations. A comparison of the subset PALM 'other context' with the other lexical sets of this cluster seemed inappropriate due to the disproportionate number of tokens. Therefore, the remaining 57 PALM tokens (no further data cleansing necessary) were added to the BATH set. The values of BC = 0.92 and Pillai = 0.01 ($p < 0.001$) appeared to justify this decision.

The Pillai score and BC for (BATH/PALM)-START was BC = 0.97 and Pillai = 0.00, $p < 0.001$. The almost complete overlap suggests a common phoneme for all speakers. Because of this outcome, STRUT was compared with (BATH/PALM) and START taken together. The values for STRUT and BATH/PALM/START are BC = 0.96, Pillai = 0.04, $p < 0.001$. Here, the degree of overlap is also very high. Figure 5.13 illustrates the BCs of STRUT and BATH/PALM/START. Four speakers of the young group have small values: XYF03 (0.12), TYM07 (0.24), SYF05 (0.34) and SYF10 (0.35). They showed a more distinct pronunciation between the tense and the lax vowels than the majority of speakers.

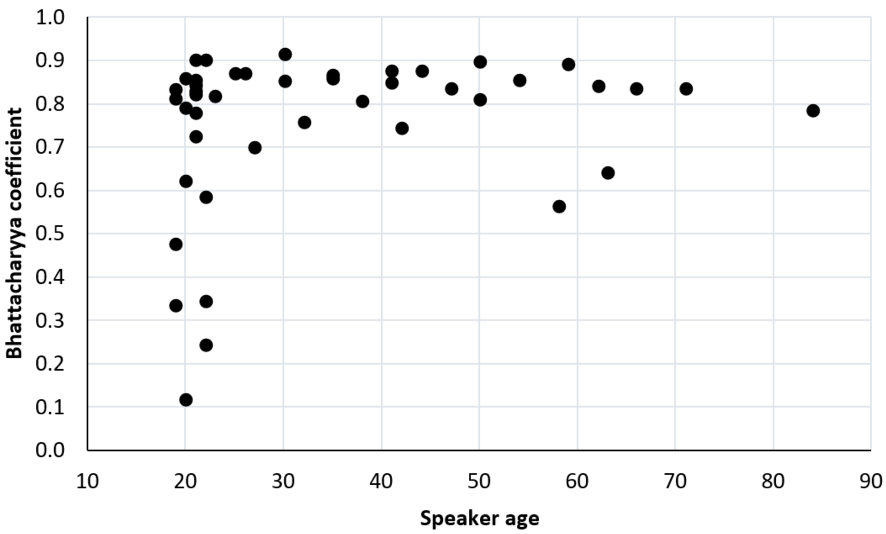

Figure 5.13: Bhattacharyya coefficients for STRUT (N = 642) and BATH/PALM/START (N = 706) in IN style separated by speaker age

Since the F1 values for all phonemes were very similar, the regression analysis was carried out only for F2. The fixed factors were *phoneme* (STRUT, BATH/PALM and START), *gender, age group, speech style, number of vowels in L1, preceding context* (coronals, non-coronals) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. The results for F2 are shown in Table 5.34. *Age group, style, preceding context* and the interaction of *phoneme*age group* generated statistically significant results. Non-coronals were produced 0.09 LU backer than coronals, an outcome that is probably not phonetically relevant because of the small distance between them. The same applies to the style variants: RP and WL were produced slightly backer than IN. ReP and WL were not significantly different from one another. The variable *phoneme* did not show statistical significance (p = 0.32). The R^2c value of 0.49 is moderately high and can be explained by 16% of the fixed effects and 33% of the random effects.

Table 5.34: F2 regression results for STRUT (N = 772), BATH/PALM (N = 393) and START (N = 547) (ph = phoneme, pc = preceding context, ag = age group)

```
Scaled residuals:
                      Min        1Q     Median      3Q       Max
                    -3.6510   -0.6437   0.0201   0.6035   3.6066
Random effects:
 Groups             Variance   Std.Dev.
 word label         0.0227     0.1515
 speaker name       0.0185     0.1360
 Residual           0.0629     0.2508
Fixed effects:
                    Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)         -0.2623    0.0545   169.90   -4.816   3.22e-06 ***
 ph start             0.1064    0.0466   203.80    2.286      0.023 *
 ph strut             0.1038    0.0456   183.30    2.277      0.024 *
```

```
                       Estimate  Std. Error    df     t value   Pr(>|t|)
 ag young              -0.2417    0.0534     68.20   -4.525  2.48e-05 ***
 pc non-coronal        -0.0905    0.0278    345.90   -3.255     0.001 **
 style ReP             -0.0908    0.0288   1221.00   -3.150     0.002 **
 style WL              -0.1516    0.0284    555.00   -5.346  1.32e-07 ***
 ph strut:ag young      0.2483    0.0354   1581.30    7.014  3.41e-12 ***

R²m (fixed effects) = 0.16
R²c (fixed + random effects) = 0.49

post-hoc Tukey, confidence level 0.95:
                        Estimate   Std. Error   z value   Pr(>|z|)
 ReP - IN == 0          -0.0908     0.0288      -3.150     0.005 **
 WL - IN == 0           -0.1516     0.0284      -5.346    <0.001 ***

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

Table 5.35 shows all significant comparisons. The figures highlighted in grey are the differences within the age groups. Again, it is the young group that behaved differently from the others. Four details are worth mentioning: First, the young speakers do not differ in their pronunciation of BATH/PALM and START. Second, their pronunciation of BATH/PALM and START differs significantly from that of STRUT. Third, they also differ from all other combinations except STRUT*older. Fourth, the middle and the older age group do not show differences within their groups as well as between them. This implies that in terms of frontness, the three phonemes were produced very similarly or identically by the middle and older age group members, and the young group produced a lax and a tense vowel.

Table 5.35: Pairwise comparison of F2 (LU) for the interaction of phoneme*age group for BATH/PALM (B/P), START and STRUT

Interaction	B/P* young	B/P* mid.	B/P* older	START* young	START* mid.	START* older	STRUT* young	STRUT* mid.
B/P*middle	0.24							
B/P*older	0.26	-						
START*young	-	-	-					
START*middle	0.35	-	-	0.21				
START*older	0.41	-	-	0.28	-			
STRUT*young	0.35	-	-	0.21	-	-		
STRUT*middle	0.34	-	-	0.21	-	-	-	
STRUT*older	0.34	-	-	-	-	-	-	-

5.2.3.2 Vowel duration

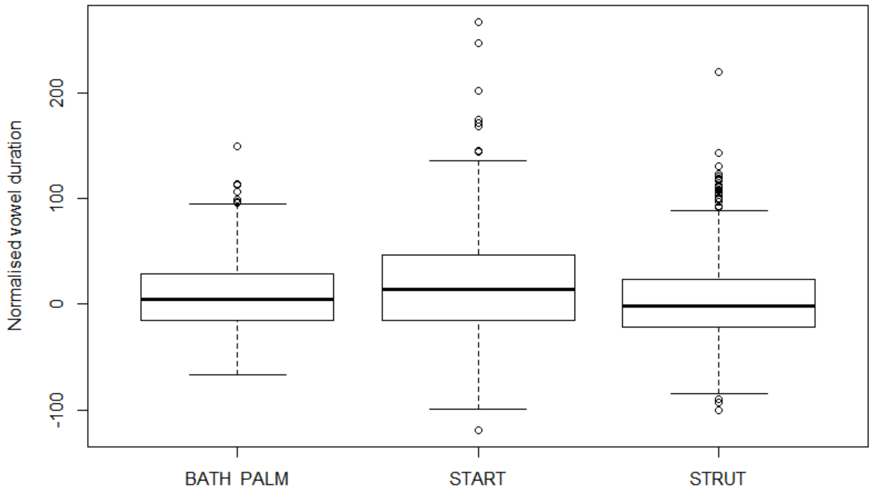

Figure 5.14:. Normalised vowel duration (ms) of the low central cluster

Figure 5.14 shows the normalised vowel duration of the low central cluster. The tense vowels do not contain word-final tokens followed by a pause. As practiced in the analysis of vowel quality, the phonemes BATH and PALM were summarised. START is the longest phoneme with the upper quartile ending at around 140 ms. The longest tokens of BATH/PALM and STRUT are considerably shorter. What is more, START also contains tokens that are the shortest of this cluster with the lowest score at about -100 ms, hence, START shows the greatest variation. The phoneme BATH/PALM has the shortest boxplot, i.e. the token distribution is smaller than those of START and STRUT. Finally, STRUT produced the highest number of outliers, especially at the upper end of the boxplot, which means that some STRUT tokens were produced comparatively long. The results of the graph above is also reflected in Table 5.36 showing, for example, that START produced the highest mean and a very large SD.

Table 5.36: Token numbers, normalised vowel durations and SDs of the low central cluster

Phoneme	N	Mean (ms)	SD
BATH/PALM	393	9.50	33.47
START	530	20.08	50.05
STRUT	772	3.89	38.31

The fixed factors entering the model were *phoneme, gender, age group, speech style, following voicing* (voiced, voiceless, pause/gap) and *following manner* (fricative, affricate, plosive, pause/gap). Furthermore, the variable *spelling* was employed. The variant <a> stands for

BATH/PALM/START and the variants <o>, <oe>, <ou> and <u> for STRUT. The interactions were age *group*gender*, *phoneme*gender*, *phoneme*age group* and *phoneme*speech style*. The variables *phoneme, speech style, following manner* and the interaction of *phoneme*age group* yielded statistically significant results (see Table 5.37). Other variables, like *following voicing* or *spelling* were not significant. The post-hoc test for *speech style* revealed that across all phonemes, tokens in WL style were pronounced longer than in IN style. The post-hoc test for *following manner* showed that vowels followed by plosives were produced shorter than vowels followed by fricatives. Again, this result is confirmed by Peterson & Lehiste (1960: 702). However, they also report that affricates and plosives affect the duration of preceding vowels in the same way. Interestingly, this is not the case with the present data where vowels before plosives were produced significantly shorter than before affricates. Peterson & Lehiste (1960) assume a historical basis for their findings, which might be inexistent for the non-native speakers of the present study.

Table 5.37: Regression results of duration for BATH/PALM (N = 393), START (N = 530) and STRUT (N = 772) (ph = phoneme)

```
Scaled residuals:
                      Min       1Q      Median     3Q       Max
                   -3.2970   -0.5937   -0.0606   0.5038   6.2455
Random effects:
 Groups             Variance   Std.Dev.
 word label         668.995    25.865
 speaker name         4.569     2.137
 Residual           946.119    30.759
Fixed effects:
                     Estimate Std. Error   df     t value   Pr(>|t|)
 (Intercept)           18.556  10.734     264.70    1.729     0.085
 ph start              23.590   7.990     302.50    2.953     0.003 **
 style WL              13.306   3.752    1313.50    3.546     0.000 ***
 foll mann. plosive   -26.501   9.222     257.60   -2.874     0.004 **
 ph strut:ag young    -11.177   4.348    1578.60   -2.571     0.010 *

R²m (fixed effects) = 0.08
R²c (fixed + random effects) = 0.46

Post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 WL - IN == 0           13.306     3.752        3.546     0.001 **

 plosive - affric == 0 -26.757     9.222       -1.137     0.017 *
 plosive - fricat == 0 -16.422     5.752       -2.855     0.018 *

Significance codes : p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

Table 5.38 shows the pairwise comparison of the three lexical sets. In general, STRUT is pronounced shorter than START. This is particularly obvious in the young and middle age group where the phonemes also differ within their respective groups. In the older group, the difference between STRUT and START is 20.58 ms, but this result is statistically not significant (p = 0.1. Therefore, it is not listed in the table). Another outcome is interesting: There were no significant differences between STRUT and BATH/PALM. STRUT is on average 5.70 ms shorter than BATH/PALM (SD = 3.96 ms). In particular, it means a difference of 12.5 ms for the young age group, 1.3 ms for the middle group and 1.1 ms for the older group. Moreover, there is a difference in vowel length between BATH/PALM and START. BATH/PALM is on average 20.32 ms shorter than START (SD = 4.83). This outcome also applies to all age groups, but it is not statistically significant with p-values between 0.06 and 0.9.

Table 5.38: Pairwise comparison of duration (ms) for the interaction of phoneme*age group for BATH/PALM (B/P), START and STRUT

Interaction	B/P* young	B/P* mid.	B/P* older	START* young	START* mid.	START* older	STRUT* young	STRUT* mid.
B/P*middle	-							
B/P*older	-	-						
START*young	-	-	-					
START*middle	-	-	-	-				
START*older	-	-	-	-	-			
STRUT*young	-	-	-	-30.41	-30.13	-23.90		
STRUT*middle	-	-	-	-25.20	-24.91	-	-	
STRUT*older	-	-	-	-27.10	-26.82	-	-	-

5.2.3.3 STRUT spelling

The STRUT vowel can be orthographically represented by five spellings: <u> (e.g. *but*), <o> (e.g. *some*), <ou> (e.g. *cousin*), <oo> (e.g. *blood*) and <oe> (e.g. *does*). It was tested whether spelling has an influence on the formants. The number of tokens by spelling in the present study is oo = 1, oe = 16, ou = 54, u = 609 and o = 93. The spelling variant <oo> was excluded because it occurred only once. The factors *age group, gender, spelling, spelling*gender* and *spelling*age group* entered the model for vowel quality. The run for F1 did not yield significant results. The p-values for any variable or interaction were higher than 0.10. For F2, only the interaction of *spelling*age group* yielded one significant difference (Table 5.39). A pairwise comparison, however, brought no further significances. The p-values in the post-hoc test ranged between 0.86 and 1.00. In sum, the spelling variants analysed had no impact on the pronunciation of STRUT.

Table 5.39: F2 regression results for STRUT (N = 772) (sp = spelling, ag = age group)

```
Scaled residuals:
                      Min       1Q     Median     3Q       Max
                   -4.3147   -0.5850   0.0155   0.6205   3.3089
Random effects:
 Groups             Variance   Std.Dev.
 word label         0.0252     0.1588
 speaker name       0.0147     0.1211
 Residual           0.0524     0.2290
Fixed effects:
                    Estimate  Std. Error   df    t value   Pr(>|t|)
 (Intercept)         -0.1904    0.0678   153.40   -2.807     0.006 **
 sp u:ag older        0.2241    0.0859   722.20    2.610     0.009 **

R²m (fixed effects) = 0.01
R²c (fixed + random effects) = 0.44

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

5.2.3.4 Summary of the key findings

Regarding vowel quality, the variables *age group, speech style* and *preceding context* and the interaction of *phoneme*age group* were decisive. Although *speech style* and *preceding context* yielded significant differences, they were so small that they are probably phonetically irrelevant. Regarding the interaction, it is again the young group that behaves differently. In contrast to speakers of the middle and older group, young speakers differentiated between the lax vowel STRUT and the tense vowels BATH/PALM and START, which they produced as one phoneme. Tokens in the PALM set with postvocalic /m/ were generally produced lower and backer than the other tokens of this set. Concerning the particular quality of the vowels, the middle and older age group may have one phoneme, the open back unrounded /ɑ/ for the lax vowel STRUT and the tense vowels BATH, START and PALM. The young speakers are assumed to have two phonemes for this cluster: /ʌ/ for STRUT and a slightly raised /ɑ/ for BATH, START and PALM.

Regarding vowel duration, the variables *phoneme, speech style* and *following manner* and the interaction of *phoneme*age group* were statistically significant. There are three results that apply to all age groups although not all were statistically confirmed:

- START was pronounced longer than STRUT

- START was pronounced longer than BATH/PALM
- STRUT and BATH/PALM did not differ in their vowel duration.

What is more, tokens in WL style were produced longer than in IN style. Vowels before plosives were produced shorter than before affricates and fricatives. The spelling variants of STRUT had no influence on pronunciation whatsoever. The voicing of the postvocalic consonants did not play a role.

5.2.4 The mid back vowels

5.2.4.1 Vowel quality

In Wells' (1982a) taxonomy, the phonemes LOT, CLOTH, THOUGHT, FORCE and NORTH are five different lexical sets. In Received Pronunciation (RP), LOT and CLOTH are realised as /ɒ/ and THOUGHT, FORCE and NORTH as /ɔ/ (Cruttenden 2008: 120). In most African varieties including WSAE, all five lexical sets are reported as being produced as /ɔ/ (Bowerman 2004; Gut 2004, 2004; Huber 2004; Mesthrie 2004b; Van Rooy 2004). The analysis of this cluster is an iterative process. For a better visualisation of the phoneme distributions, it is first separated into a lax subcluster (LOT-CLOTH) and a tense one (THOUGHT-NORTH-FORCE). Both subclusters are first described individually and will later be compared. Figure 5.15 shows the LOT-CLOTH subcluster. Both means lie very close together, and the SDs describe an almost complete overlap of both phonemes for formant values (see also Table 5.40). The BC of 0.96 and the Pillai score of 0.02 with $p < 0.001$ suggest full coalescence. For further investigations, both phonemes will therefore be considered as one: LOT/CLOTH.

Table 5.40: Token numbers, formant means and SDs of LOT and CLOTH

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
LOT	986	0.46	0.50	-0.97	0.30
CLOTH	291	0.31	0.48	-1.06	0.27

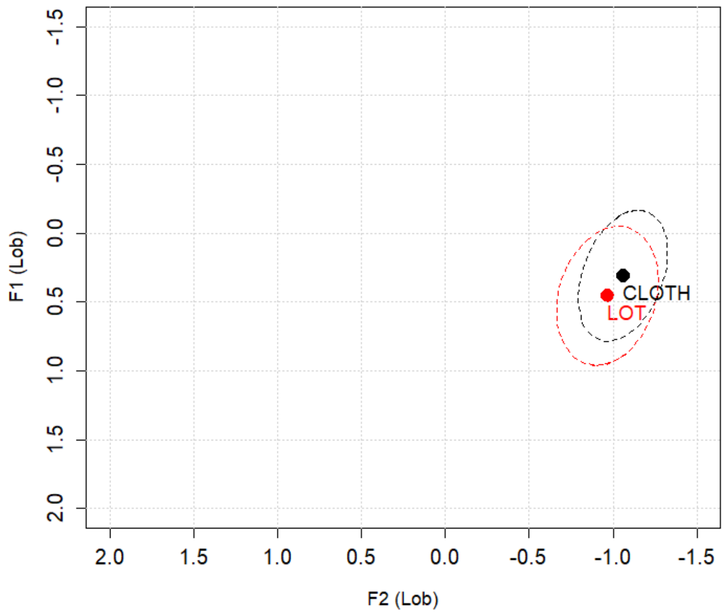

Figure 5.15: LOT-CLOTH **subcluster across all variables with 1 SD**

The tense subcluster (Figure 5.16 and Table 5.41) shows that the means and spread of THOUGHT and NORTH are practically identical. Both phonemes can be regarded as one. In visible contrast, FORCE is produced higher and much more retracted than the former. This outcome is backed by Brato (Manuscript), who described exactly the same feature for Botswana-English.

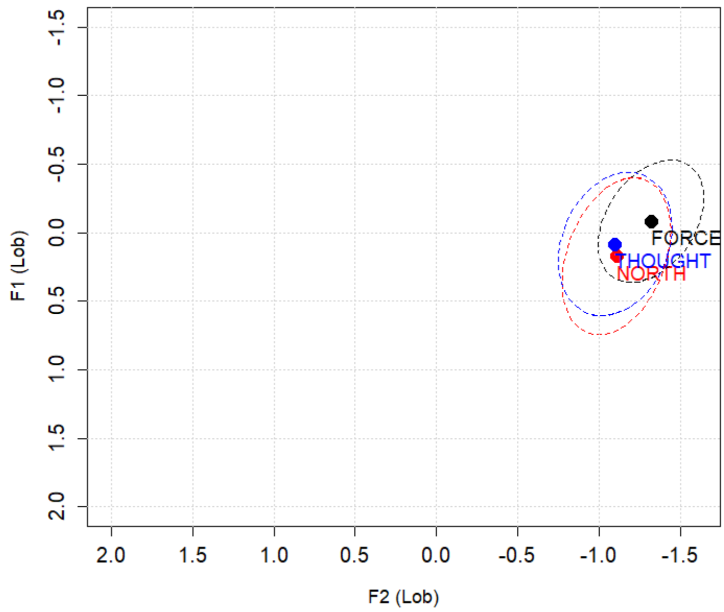

Figure 5.16: THOUGHT-NORTH-FORCE **subcluster across all variables with 1 SD**

Again, many English varieties are reported to produce these three phonemes identically, which is why it was investigated why the present data did not show this outcome.

Table 5.41: Token numbers, formant means and SDs of THOUGHT, NORTH and FORCE

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
THOUGHT	480	0.08	0.52	-1.10	0.35
NORTH	416	0.17	0.57	-1.11	0.33
FORCE	334	-0.09	0.44	-1.32	0.32

The lexical sets THOUGHT and NORTH included 610 tokens in inter-consonantal position (e.g. *daughter*, *short*) and one token in word-initial position (*awesome*). FORCE words comprised tokens in inter-consonantal position (e.g. *pork, support*) and word-final position (e.g. *more, door, four*). Since the composition of the phonemes included different phonological contexts, it was checked whether the vowel position in the word is responsible for the distribution in the vowel space. Once more, a contour plot can provide a more fine-grained distribution than means and SDs.

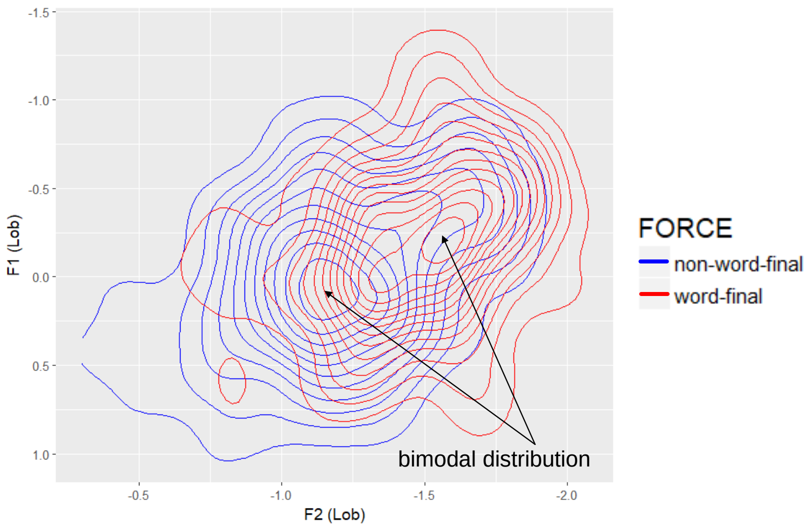

Figure 5.17: Contour plot of FORCE subsets across all speakers and speech styles

The contour plot for FORCE yielded a bimodal distribution with two peaks as can be seen in Figure 5.17. The overlap is high, but word-final tokens are generally produced more retracted and somewhat higher than non-final ones. Consequently, the results for the subset FORCE "non-word-final" corresponds with THOUGHT and NORTH whereas the subset FORCE "word-final" does not. Since the token number of both subsets is comparable (see Table 5.42), it can be assumed that this outcome is systematic. At this point, it is worth mentioning that the tense cluster does not contain THOUGHT and NORTH tokens in word-final position.

BC and Pillai were calculated for THOUGHT and NORTH: BC = 0.96 und Pillai = 0.01 (p = 0.03). The result hints at one single phoneme. In a next step, BC and Pillai were calculated for the subclusters (LOT/CLOTH) and (THOUGHT/NORTH/FORCE "non-word-final"). With BC = 0.92 and Pillai = 0.10, p < 0.001, both subclusters also overlap to a high degree. Figure 5.18 shows the BCs for the latter in interview style. The calculation of the coefficient requires a minimum of five tokens for both phonemes. Speaker SYF09 did not produce enough processible tokens in interview style, for which reason she had to be omitted.

Table 5.42: Token numbers, formant means and SDs of FORCE subsets

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
THOUGHT	480	0.08	0.52	-1.10	0.35
NORTH	416	0.17	0.57	-1.11	0.33
FORCE	334	-0.09	0.44	-1.32	0.32

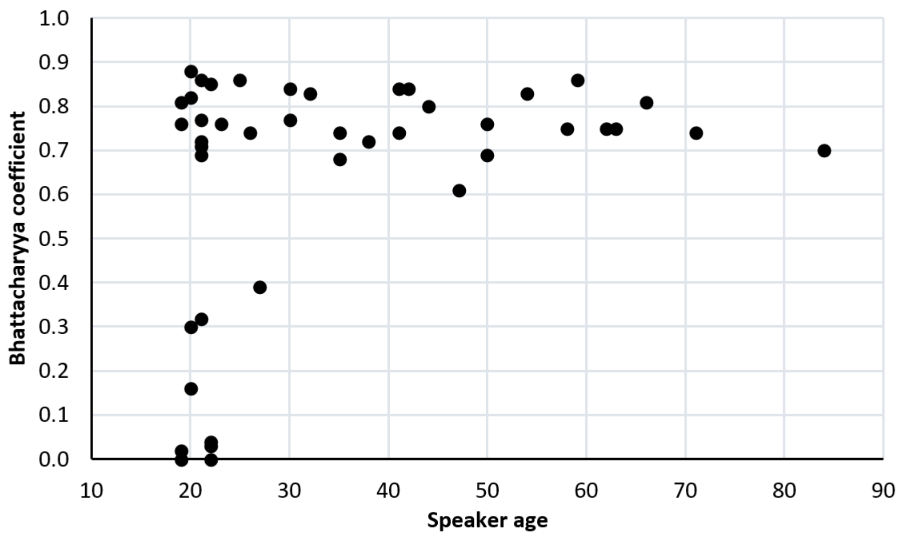

Figure 5.18: Bhattacharyya coefficients for LOT/CLOTH (N = 554) and THOUGHT/NORTH/FORCE non-word-final (N = 638) in IN style separated by speaker age

Five young speakers clearly distinguished between the lax and the tense subcluster. These were TYF04 (0.00), ZYM09 (0.00), SYF05 (0.02), SYF10 (0.03) and TYM07 (0.04). Speakers ZYM04 (0.16), XYF03 (0.30), XYM01 (0.31) and ZMF12 (0.39) show some overlap, but visibly less than the majority of speakers, who did not differentiate between both subclusters.

In analogy to LOT/CLOTH, the phonemes THOUGHT and NORTH are subsumed for further investigation, and FORCE are divided in the variants FORCE "non-word-final" and FORCE "word-final". The regression analysis was carried out for F1 and F2. The fixed factors for both runs were phoneme (LOT/CLOTH = LC, THOUGHT/NORTH = TN, FORCE "non-word-final" = F nwf and FORCE "word-final" = F wf), *gender, age group, speech style, number of vowels in L1, preceding context* (coronals, non-coronals) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. The variables that yielded statistically significant results for F1 are shown in Table 5.43. These are the variables *phoneme, speech style* and the interaction of *phoneme*gender* and *phoneme*age group*. R^2 values are moderately high with 0.26 and 0.41 respectively. Regarding *speech style*, the post-hoc test yielded significant differences between all variants; vowel height is highest in IN, followed by ReP.

Table 5.43: F1 regression results for LOT/CLOTH (N = 669), THOUGHT/NORTH (N = 611), FORCE non-word-final (N = 184) and FORCE word-final (N = 131) (ph = phoneme, ag = age group)

```
Scaled residuals:
                      Min       1Q      Median     3Q       Max
                   -3.2397   -0.6185   -0.0221   0.6712   4.6683
Random effects:
 Groups              Variance   Std.Dev.
 word label          0.0294     0.1716
 speaker name        0.0156     0.1247
 Residual            0.1778     0.4217

Fixed effects:
                     Estimate  Std. Error   df     t value   Pr(>|t|)
 (Intercept)         -0.1165    0.0849   332.50   -1.373     0.171
 ph lot/cloth         0.3850    0.0873   468.30    4.408   1.29e-05 ***
 style ReP            0.1642    0.0384   687.60    4.277   2.16e-05 ***
 style WL             0.4228    0.0414   411.90   10.201    < 2e-16 ***
 ph LC:gender male   -0.1911    0.0755  1540.00   -2.530     0.012 *
 ph LC:ag older      -0.2139    0.1105  1523.00   -1.935     0.053 .
 ph LC:ag young       0.3322    0.0812  1546.00    4.092   4.50e-05 ***

R²m (fixed effects) = 0.26
R²c (fixed + random effects) = 0.41

Post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 ReP - IN == 0          0.1642     0.0384      4.277    <1e-04 ***
 WL - IN == 0           0.4228     0.0414     10.201    <1e-04 ***
 WL - ReP == 0          0.2586     0.0473      5.467    <1e-04 ***

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

Table 5.44 shows the matrix of all statistically significant interactions in F1, which applies only to 15 out of 66 possible combinations. The table is thus truncated and shows mainly cells containing figures. The figures highlighted in grey are the differences within the age groups.

What is obvious is that the differences are between the lax and the tense phonemes, but not within the tense subsets. It is mainly the young group that differs from the other age groups, followed by the middle group. Older speakers did not differentiate between lax and tense vowels. Their estimates ranged between 0.06 and 0.15 with $p > 0.9$ (figures not given in the table). LC*young is significantly different from most of the other combinations, i.e. vowel height is highest. LC*young also has the greatest distance to the FORCE variants and to THOUGHT/NORTH within this age group. Within the middle group, significant differences are only between the lax vowel and the FORCE variants. The estimate for the comparison LC*middle with TN*middle was 0.16 with $p = 0.21$; figure not given in the table.

Table 5.44: Pairwise comparison of F1 (LU) for the interaction of phoneme*age group for LOT/CLOTH, THOUGHT/NORTH, FORCE nwf and FORCE wf

Interaction	LC*young	LC*middle	LC*older	...
LOT/CLOTH*middle	-0.22			
LOT/CLOTH*older	-0.43	-		
THOUGHT/NORTH*young	-0.55	-0.33	-	
THOUGHT/NORTH*mid.	-0.38	-	-	-
THOUGHT/NORTH*older	-0.41	-	-	-
FORCE nwf*young	-0.62	-0.40	-	-
FORCE nwf*middle	-0.51	-0.29	-	-
FORCE nwf*older	-	-	-	-
FORCE wf*young	-0.78	-0.56	-	-
FORCE wf*middle	-0.64	-0.42	-	-
FORCE wf*older	-0.56	-	-	-

The comparison of the genders is illustrated in Table 5.45. Females systematically differentiates between lax and tense vowels in that they produced tense vowels higher than the lax vowel within their group. Men, in contrast, do not show statistically significant differences between these vowel types with one exception: FORCE in word-final position is pronounced

0.32 LU higher than LOT/CLOTH. Male and female speakers do not differentiate in the pronunciation of THOUGHT/NORTH and of FORCE in non-word-final position. However, they do in the pronunciation of FORCE in word-final position although this result is not statistically significant. Here, women produced a 0.18 LU higher FORCE vowel ($p = 0.43$; figure not given in the table). Finally, there is no difference between men and women in pronouncing lax vowels.

Table 5.45: Pairwise comparison of F1 (LU) for the interaction of phoneme*gender for LOT/CLOTH, THOUGHT/NORTH, FORCE nwf and FORCE wf

Interaction	LC* male	LC* female	TN* male	TN* female	F nwf* male	F nwf* female	F wf* male
LC*female	-						
TN*male	-	-0.26					
TN*female	-	-0.31	-				
F nwf*male	-	-0.34	-	-			
F nwf*female	-0.32	-0.42	-	-	-		
F wf*male	-	-0.41	-	-	-	-	
F wf*female	-0.48	-0.59	-	-	-	-	-

The regression results for F2 are seen in Table 5.46. *Phoneme, age group, speech style* and the interaction of *phoneme*age group* are the variables and interaction respectively that yielded statistically significant results. The R^2m value is moderately high. The fixed effects that entered the model have comparably little impact in contrast to the random effects, which accounts for 34% of the outcome. The post-hoc test for *speech style* shows that WL significantly differs from ReP and IN in that phonemes in citation style were produced more retracted than those in the other styles.

Table 5.46: F2 regression results for LOT/CLOTH (N = 669), THOUGHT/NORTH (N = 611), FORCE nwf (N = 184) and FORCE wf (N = 131) (ph = phoneme, ag = age group)

```
Scaled residuals:
                        Min        1Q      Median      3Q        Max
                     -3.7328   -0.5902   -0.0307   0.5706    5.4347
Random effects:
 Groups              Variance   Std.Dev.
 word label          0.0292     0.1710
 speaker name        0.0151     0.1231
 Residual            0.0630     0.2510
Fixed effects:
                     Estimate  Std. Error   df     t value    Pr(>|t|)
 (Intercept)          -1.1800    0.0551   219.70  -21.388    < 2e-16 ***
 ph lot/cloth          0.1457    0.0542   269.30    2.689      0.008 **
 ph thought/north      0.1335    0.0560   248.40    2.383      0.018 *
 ph force wf          -0.2061    0.0941   175.00   -2.190      0.030 *
 style WL             -0.1787    0.0724   130.50   -6.561    7.7e-11 ***
 ph LC:ag young        0.1446    0.0491  1541.00    2.947      0.003 **

R²m (fixed effects) = 0.17
R²c (fixed + random effects) = 0.51

Post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 WL - IN == 0         -0.1787      0.0272      -6.561    <1e-04 ***
 WL - ReP == 0        -0.1358      0.0300      -4.528    <1e-04 ***

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

The pairwise comparison of the age groups are shown in Table 5.47. Since by far not all constellations yielded statistically significant results, this table is also truncated. The young group displays the largest F2 distance between the phonemes; all tense vowels were produced significantly backer than the lax vowel. In all age groups, FORCE in word-final position is produced backer than THOUGHT/NORTH. No significant differences could be found between THOUGHT/NORTH and FORCE nwf and between FORCE nwf and FORCE wf. The comparison of the tense subsets, however, showed that the F2 mean of FORCE nwf is located between THOUGHT/NORTH and FORCE wf, but closer to THOUGHT/NORTH. THOUGHT/NORTH is produced fronter than FORCE in non-final position with values between 0.11 LU in young, 0.13 LU in middle and 0.19 LU in older; $p > 0.2$), which assumes that THOUGHT/NORTH and FORCE nwf can be considered as one phoneme at least in the young and middle group (provided the p-value is disregarded). The F2 distance between FORCE wf and FORCE nwf is not statistically significant either ($p > 0.07$), but with values of -0.29 LU in young, -0.25 LU in middle and -0.20 LU in older, it is safe to say that word-final FORCE is pronounced far more retracted than FORCE in word-final position. It seems that FORCE in word-final position has an impact on pronunciation in all age groups although this outcome lacks statistical proof.

Table 5.47: Pairwise comparison of F2 (LU) for the interaction of phoneme*age group for LOT/CLOTH, THOUGHT/NORTH, FORCE nwf and FORCE wf

Interaction	LC* young	LC* middle	LC* older	TN* young	TN* middle	TN* older	...
LC*middle	-						
LC*older	-	-					
TN*young	-0.18	-	-				
TN*middle	-	-	-	-			
TN*older	-	-	-	-	-		
F nwf*young	-0.29	-	-	-	-	-	
F nwf*middle	-0.28	-	-	-	-	-	-
F nwf*older	-0.33	-	-	-	-	-	-
F wf*young	-0.58	0.45	-0.39	-0.40	-0.43	-0.43	-
F wf*middle	-0.48	-0.35	-	-	-0.34	-	-
F wf*older	-0.58	0.45	-	-0.40	-0.44	-0.44	-

5.2.4.2 Vowel duration

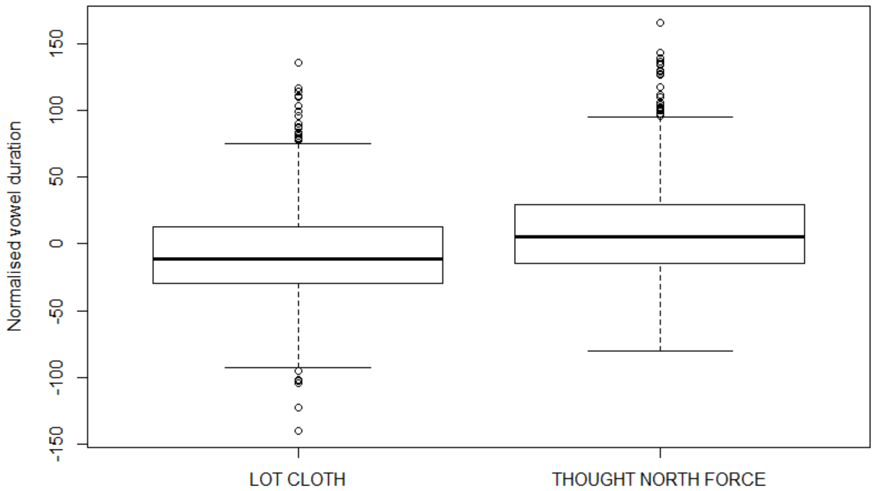

Figure 5.19: Normalised vowel duration (ms) of the mid back cluster

Figure 5.19 shows the normalised vowel duration of the mid back cluster. The tense vowel does not contain word-final tokens. As practiced in the analysis of vowel quality, the lax subcluster is summarised to LOT/CLOTH and the tense subcluster to THOUGHT/NORTH/FORCE. The boxed of both phonemes are very similar in both position and dimension. Still, the tense vowel is produced higher than the lax vowel. Noticeable are the high number of outliers at the upper end of both plots. The standard deviation 1 SD is almost identical for both phonemes as seen in Table 5.48.

Table 5.48: Token numbers, normalised vowel durations and SDs of the mid back cluster

Phoneme	N	Mean (ms)	SD
LOT/CLOTH (LC)	669	-6.34	36.72
THOUGHT/NORTH/FORCE (TNF)	795	10.69	36.59

The fixed factors entering the model were *phoneme, gender, age group, speech style, following voicing* (voiced, voiceless, pause/gap) and *following manner* (fricative, affricate, plosive, pause/gap). The interactions were *age group*gender, phoneme*gender, phoneme*age group* and *phoneme*speech style*. The variables *phoneme, speech style, following voicing* and only the interaction of *phoneme*speech style* yielded statistically significant results (see Table 5.49).

Table 5.49: Regression results of duration for LC (N = 669) and TNF (N = 795)

```
Scaled residuals:     Min       1Q     Median     3Q       Max
                   -3.4013  -0.6303  -0.0984   0.4694   4.3810
Random effects:
 Groups             Variance      Std.Dev.
 word label         447.60        21.16
 speaker name         0.00         0.00
 Residual           857.90        29.29

Fixed effects:
                      Estimate Std. Error   df     t value   Pr(>|t|)
 (Intercept)            43.284  23.546   1454.90    1.838     0.066
 ph TNF                 19.666   4.180    200.50    4.705   4.73e-06 ***
 foll voiceless        -16.071   4.618   1329.70   -3.480     0.001 **
 foll mann. fricat      23.780   9.504    181.20    2.502     0.013 *
 foll manner plosive    22.101   9.240    186.50    2.392     0.018 *
 ph TNF:style WL        15.479   6.872   1152.30    2.252     0.024 *

R²m (fixed effects) = 0.14
R²c (fixed + random effects) = 0.43

Post-hoc Tukey, confidence level 0.95:
                          Estimate   Std. Error   z value   Pr(>|z|)
fricat - affric == 0       26.395      9.804        2.692     0.028 *
plosive - affric == 0      24.218      9.534        2.540     0.043 *

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

Of the variables not included in interactions, voicing and manner of the following consonants influenced vowel duration: Vowels followed by voiceless consonants were produced on average 16 ms shorter than vowels followed by voiced consonants. This result is in line with former findings (e.g. Peterson & Lehiste 1960; Thomas 2011). Fricatives and plosives were produced 26 ms and 24 ms longer than affricates. However, there is no difference in vowel duration between fricatives and plosives. The interaction of *phoneme*speech style* is shown in Table 5.50. The figures highlighted in grey are the differences within the speech styles. Within each speech style, tense vowels were produced longer than lax vowels, which is particularly obvious in WL style where the difference is 35.67 ms. It seems that also in this vowel cluster, speakers took particular care in the pronunciation of words in citation form. Lax and tense vowels across the speech styles did not differ.

Table 5.50: Pairwise comparison of duration (ms) for the interaction of phoneme*speech style for LC and TNF

Interaction	LC*WL	LC*ReP	LC*IN	TNF*WL	TNF*ReP
LC*ReP	-				
LC*IN	-	-			
TNF*WL	35.67	31.67	26.68		
TNF*ReP	27.12	23.13	-	-	
TNF*IN	30.19	26.19	21.21	-	-

5.2.4.3 Summary of the key findings

Regarding vowel quality, the variables *phoneme, age group, speech style* and the interaction of *phoneme*gender* and *phoneme*age group* yielded statistically significant results. In terms of age group, young speakers consistently differentiate between lax and tense vowels in both vowel height and vowel frontness. The quality may be an open rounded back vowel /ɒ/ for LOT/CLOTH and an open-mid rounded back vowel /ɔ/ for THOUGHT, NORTH and FORCE non-word-final. The middle and older group show little or no differences in vowel height and frontness between LOT/CLOTH, THOUGHT/NORTH and FORCE in non-word-final position, suggesting one phoneme for this cluster. These speakers may have produced an open-mid back vowel /ɔ/. The word-final subset of FORCE, is produced backer than THOUGHT/NORTH in all three age groups. This is even the case in the comparison FORCE in non-word-final position although the results were not statistically significant. It can be thus tentatively concluded that vowels in word-final position triggers a backer pronunciation compared to vowels in inter-consonantal position. Whether or not other phonological contexts may have played a role in this outcome could not be traced. Concerning the genders, females produced lax and tense vowels at different vowel heights while males did not. In this respect, one is inclined to say that women are more attentive to speech than men, which is frequently reported in the literature (e.g. Dubois & Horvath 2000; Labov 1991; Lanham & Macdonald 1985). Vowel frontness, however, did not yield statistically significant differences between male and female speakers. Therefore, whether men and women pronounce tense and lax vowels differently indeed cannot be answered conclusively with the present results. With regard to speech style, tokens on WL style were produced more retracted than those in the other styles.

Regarding vowel duration, the variables *phoneme, speech style, following voicing* and the interaction of *phoneme*speech style* influenced the length of the vowels. The tense vowels included in TNF was generally produced longer than the lax counterpart. Voicing and manner of the following consonants influenced vowel duration: Vowels preceding voiceless consonants were produced shorter than vowels preceding voiced consonants. Fricatives and plosives were produced longer than affricates. Difference in vowel duration between fricatives and plosives could not be observed. Regarding speech style, tense vowels were produced longer than lax vowels and longest in WL style. There was no difference in vowel duration between lax and tense vowels across the speech styles. Interestingly, the variable *age group* or the interaction of *phoneme*age group* did not play a role.

5.2.5 The high back vowels

5.2.5.1 Vowel quality

Figure 5.20 shows two significantly distant vowel means. GOOSE is produced much fronter than FOOT and has an extremely large horizontal expansion with SD = 0.91 (see also Table 5.51). FOOT is produced slightly lower and has a much smaller SD than GOOSE. The SD of GOOSE covers about the half of the SD of FOOT.

GOOSE fronting has been reported for L1 as well as L2 varieties around the world. Examples of native speaker varieties are Aberdonian English (Brato 2016), Carlisle English (Jansen 2017) and English in Reading and London (Torgersen & Kerswill 2004). Examples of second language varieties are Chicano English in Los Angeles (Fought 1999), Chinese English in San Francisco (Hall-Lew 2009), and in the African context, Kenyan English (Hoffmann 2011), the

Nigerian variety of Ebira English (Isiaka 2017) and Botswanan English (Brato Manuscript). In South Africa, it can be found in L1 as well as in L2 varieties, such as WSAE, ISAE and BSAE (Mesthrie 2010b).

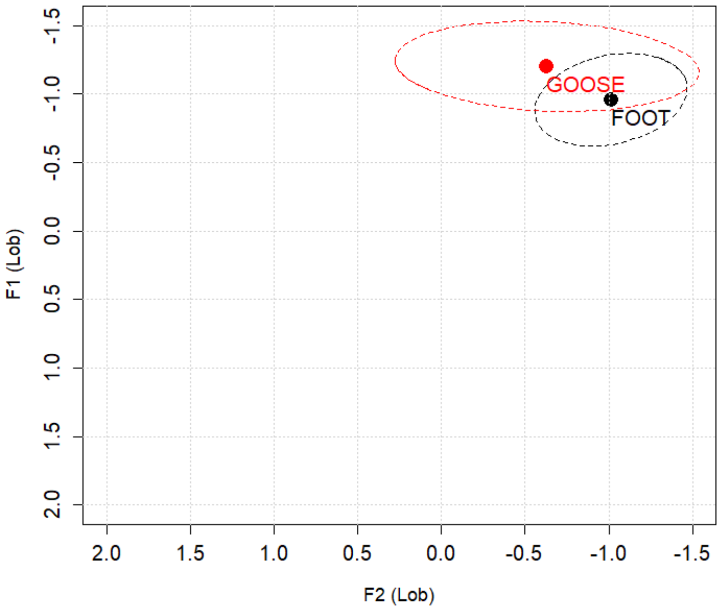

Figure 5.20: High back cluster across all variables with 1 SD

Table 5.51: Token numbers, formant means and SDs of the high back cluster

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
GOOSE	1154	-1.20	0.33	-0.63	0.91
FOOT	580	-0.96	0.34	-1.01	0.45

GOOSE-fronting is particularly triggered by preceding /j/ in words like *use*, *beautiful* and *computer* (Mesthrie 2010b). For this reason, the lexical set GOOSE was subdivided into the subset J-words (in analogy to Mesthrie's 2010b labelling), containing all GOOSE tokens with preceding /j/, and the subset goose|, containing all other GOOSE tokens. In order to distinguish the lexical set GOOSE from the subset, the latter is written in lower case and with a vertical bar at the end.

Figure 5.21 shows the contour plot of J-words, goose| and FOOT. The J-words differ strongly from goose| and FOOT in their front/back dimension. The J-words represent a central vowel but with an extreme horizontal spread (see also Table 5.52). The overlap with goose| is in the right half of the J-words distribution. It is obvious that some speakers differentiate between J-words and goose|, and some speakers do not. For the latter, GOOSE is mostly a back vowel. The overlap of goose' and FOOT is very high, suggesting that goose' and FOOT is produced as one sound by most speakers.

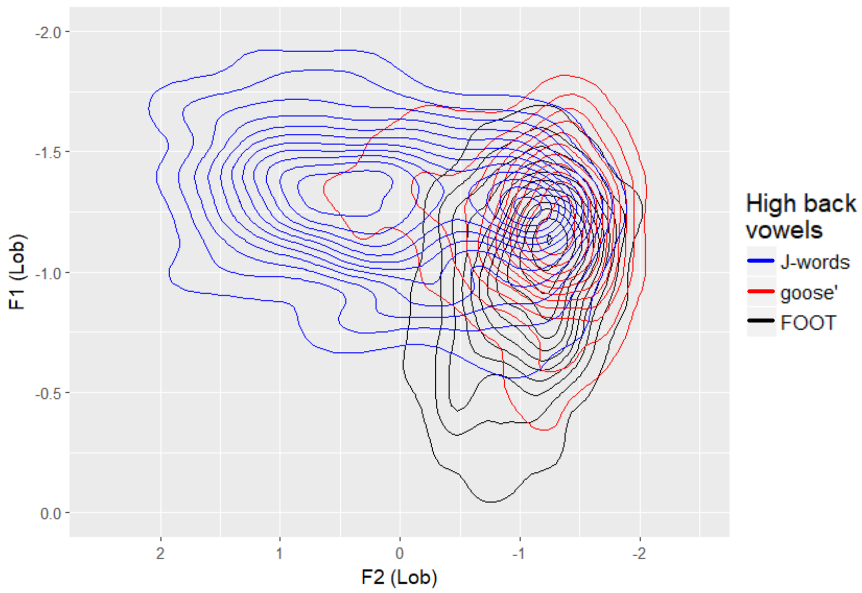

Figure 5.21: Contour plot of the high back cluster across all variables

Table 5.52: Token numbers, formant means and SDs of FOOT and the GOOSE subsets

Phoneme	N	Mean F1	SD F1	Mean F2	SD F2
goose'	718	-1.16	0.34	-1.04	0.64
J-words	436	-1.28	0.29	0.05	0.89
FOOT	580	-0.96	0.34	-1.01	0.45

The BC and Pillai scores were calculated with FOOT, goose' and J-words. The BC value for goose' and FOOT is with BC = 0.91 very high, and the Pillai score with Pillai = 0.08, $p < 0.001$ very low. The BC values for goose' and J-words is 0.76, and the Pillai score is 0.34, $p < 0.001$. Figure 5.22 shows the BCs for goose' and FOOT in interview style. Four speakers (SMM01, SYM08, XMM09 and ZYM04) did not produce enough processible FOOT tokens in interview style, so they had to be omitted. Three young speakers (ZYM09 (0.01), XYM01 (0.06) and XYF03 (0.29)) and a middle-aged speaker (SMM04 (0.21)) distinguish between both sets and show little overlap respectively. Most speakers, however, show some degree of overlap. A real pattern cannot be discerned. A comparison between male and female speakers did not bring elucidation either.

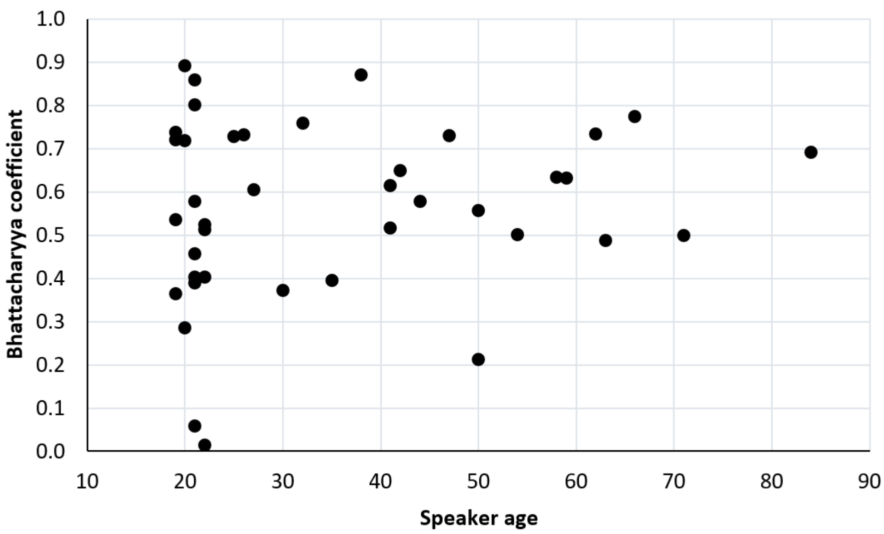

Figure 5.22: Bhattacharyya coefficients for FOOT (N = 338) and goose' (N = 340) in IN style separated by speaker

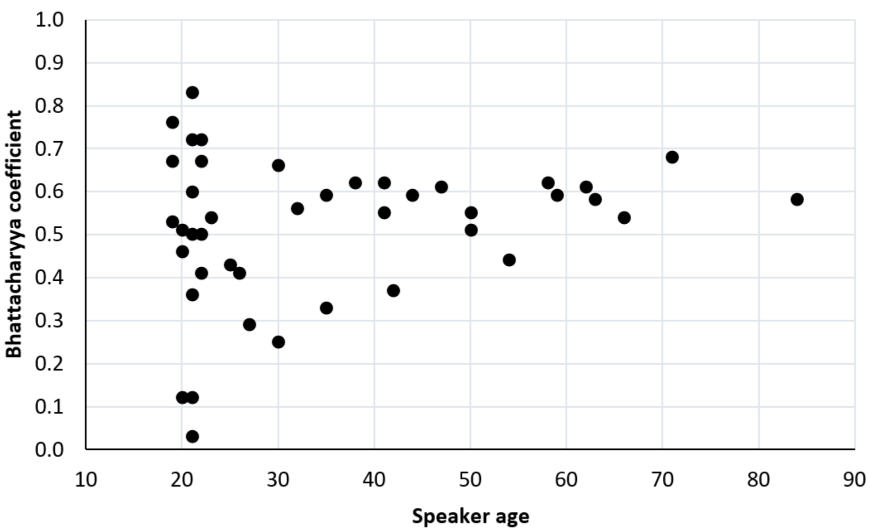

Figure 5.23: Bhattacharyya coefficients for goose' (N = 695) and J-words (N = 429) across all speech styles for 42 speakers separated by speaker age

Figure 5.23 shows the BCs for goose' and J-words. Unfortunately, 22 speakers did not reach the minimum number of five J-words tokens in IN style. Therefore, the coefficient for all speakers was calculated across all styles, but two speakers (SYF06 and ZYM04) still had to be neglected because of insufficient token numbers and non-processible data respectively. Three young speakers clearly distinguished between goose' and J-words. These were SYM03 (0.03), XYM01 (0.12) and XYF03 (0.12). None of the remaining speakers pronounce both phonemes as one. On the contrary, most speakers show an overlap in the middle range, that is, they show much variation, which is particularly evident among the young speakers.

The BC of FOOT and goose' and that of goose' and J-words show much overlap. The former outcome may result from L1 interference in that FOOT and the GOOSE allophone are produced very similarly due to the lack of tense/lax contrast in the Bantu languages. The latter outcome may result from the fact that J-words still belong to the GOOSE set. GOOSE fronting is visible (see also Figure 5.21), but this process is probably in progress rather than completed. Moreover, GOOSE fronting may be a matter of gradation, i.e. speakers front J-words, but to different degrees.

The regression analysis was first carried out for F1. The factors were *phoneme* (GOOSE, FOOT), *speech style, gender, age group, number of vowels in L1, preceding context* (coronals, non-coronals, /j/ (=J-words)) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. The variables that yielded statistically significant results for F1 are shown in Table 5.53. These are *phoneme, gender, age group* and the interaction of *phoneme*age group*. Males produce both phonemes higher than females, the value of -0.07 is certainly negligible though. R^2c value is moderately high but includes a higher influence of fixed factors (19%) compared to the remaining 13% of the random effects.

Table 5.53: F1 regression results for FOOT (N = 356) and GOOSE (N = 803) (ph = phoneme, ag = age group)

```
Scaled residuals:
                      Min       1Q      Median     3Q       Max
                  -2.6956   -0.6237   -0.0645   0.5641   4.1813
Random effects:
 Groups            Variance   Std.Dev.
 word label        0.0089     0.0943
 speaker name      0.0079     0.0891
 Residual          0.0828     0.2877

Fixed effects:
                    Estimate Std. Error    df     t value    Pr(>|t|)
 (Intercept)         -1.0343    0.0477   100.60   -21.668   < 2e-16 ***
 ph goose            -0.1213    0.0447    90.00    -2.713     0.008 **
 ag young             0.2342    0.0460   105.50     5.092  1.56e-06 ***
 gender male         -0.0722    0.0329    38.00    -2.191     0.035 *
 ph goose:ag older    0.1474    0.0546  1106.10     2.702     0.007 **
 ph goose:ag young   -0.3390    0.0416  1112.80    -8.148  8.88e-16 ***

R²m (fixed effects) = 0.19
R²c (fixed + random effects) = 0.32

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

The post-hoc pairwise comparison of *phoneme*age group* in Table 5.54 illustrates that the FOOT vowel of the young is significantly different from the other interactions. It is the lowest vowel in this cluster and has the greatest distance to GOOSE of the young speakers. The figure highlighted in grey is the difference within the age group.

Table 5.54: Pairwise comparison of F1 (LU) for the interaction of phoneme*age group for FOOT and GOOSE

Interaction	FOOT*young	FOOT*mid.	FOOT*older	GOOSE*you.	GOOSE*mid.
FOOT*middle	-0.23				
FOOT*older	-0.32	-			
GOOSE*young	-0.46	-0.22	-		
GOOSE*mid.	-0.36	-	-	-	
GOOSE*older	-0.29	-	-	0.17	-

For F2, the model included *phoneme, speech style, gender, age group, number of vowels in L1, preceding context* (coronals, non-coronals and /j/ (= J-words)) and the interactions of *age group*gender, phoneme*gender* and *phoneme*age group*. Furthermore, the interactions of *preceding context*gender* and *preceding context*age group* was added. Table 5.55 shows the outcome of the analysis for F2. The variables age group, speech style and preceding context as well as the interaction of *phoneme*gender* and *preceding context*age group* showed significant results. With the value of 0.70, R^2c is very high. The fixed effects can explain 44% of the results, and the random effects the remaining 26%.

Table 5.55: F2 regression results for FOOT (N = 356) and GOOSE (N = 803) (ph = phoneme, pc = preceding context, ag = age group, gen = gender)

```
Scaled residuals:
                     Min       1Q      Median     3Q       Max
                   -3.4330   -0.6168   -0.0456   0.5146   4.4973
Random effects:
 Groups            Variance   Std.Dev.
 word label        0.1259     0.3548
 speaker name      0.0717     0.2678
 Residual          0.2552     0.5052

Fixed effects:
                   Estimate  Std. Error   df     t value    Pr(>|t|)
 (Intercept)        -0.5667    0.1836   131.30   -3.086      0.002 **
 pc /j/              1.0210    0.1432    98.90    7.129   1.67e-10 ***
 pc non-coronal     -0.5610    0.1547   104.60   -3.627     <0.000 ***
 style WL           -0.2943    0.0650   816.60   -4.526   6.89e-06 ***
 pc /j/:ag older    -0.5288    0.1178  1064.00   -4.487   8.02e-06 ***
 ph goose:gen male -0.1784    0.0683  1071.00   -2.613      0.009 **
```

```
R²m (fixed effects) = 0.44
R²c (fixed + random effects) = 0.68

Post-hoc Tukey, confidence level 0.95:
                      Estimate   Std. Error   z value   Pr(>|z|)
 WL - IN == 0          -0.2943     0.0650      -4.526    <0.001 ***
 WL - ReP == 0         -0.1880     0.0690      -2.723     0.018 *

Significance codes: p < 0.001 ‘***’, p < 0.01 ‘**’, p ≤ 0.05 ‘*’
```

Of the variables not analysed in interactions, *speech style* influenced vowel quality: WL tokens are produced backer than IN and ReP tokens. Table 5.56 shows the pairwise comparison of *phoneme*gender*. Only two interactions are statistically significant. Male speakers produce GOOSE backer than females, and females produce FOOT backer than males produce GOOSE.

Table 5.56: Pairwise comparison of F2 (LU) for the interaction of phoneme*gender for FOOT (N = 356) and GOOSE (N = 803)

Interaction	GOOSE*female	GOOSE*male	FOOT*female
GOOSE*male	-0.28		
FOOT*female	-	-0.42	
FOOT*male	-	-	-

Table 5.57 shows the pairwise comparison of *preceding context*age group*. The figure highlighted in grey is the difference within the age group. J-words, particularly those of the young and middle group, differ generally from all other interactions: They are produced most fronted, followed by tokens with preceding coronals.

Table 5.57: Pairwise comparison of F2 (LU) for the interaction of preceding context*age group for FOOT and GOOSE (/j/ = J-words (N = 403), c = coronals (N = 286), n-c = non-coronals (N = 470))

Interaction	/j/* young	/j/* mid.	/j/* older	c* young	c* mid.	c* older	n-c* young	n-c* mid.
/j/*mid.	-							
/j/*older	-0.68	-0.57						
c*young	-0.96	-0.85	-					
c*middle	-1.13	-1.02	-	-				
c*older	-1.17	-1.06	-	-	-			

Interaction	/j/* young	/j/* mid.	/j/* older	c* young	c* mid.	c* older	n-c* young	n-c* mid.
n-c*young	-1.37	-1.26	-0.69	-	-	-		
n-c*mid.	-1.70	-1.58	-1.01	-0.74	-0.56	-	-	
n-c*older	-1.73	-1.62	-1.05	-0.77	-	-0.56	-	-

Since only the interaction of *preceding context*age group* showed statistically significant results and not the interaction of *phoneme*age group*, FOOT and GOOSE cannot be directly compared. Therefore, Figure 5.24 illustrates the constellations between FOOT and the GOOSE subsets across the age groups. Although not every distance is statistically significant, the following observations are obvious:

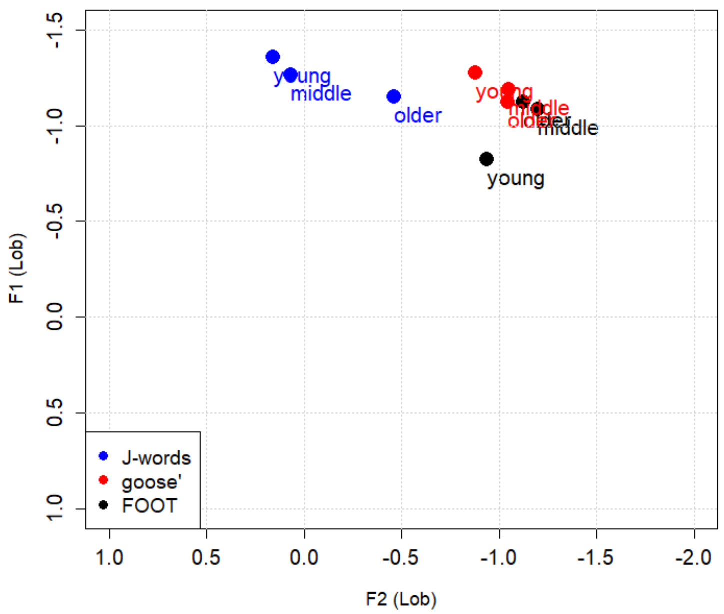

Figure 5.24: High back cluster by age group

1. The GOOSE vowel in J-words is a central vowel in the young and middle group.
2. The GOOSE vowel in J-words is a backish[12] vowel in the older speakers.
3. The young produce J-words and goose' higher than FOOT.
4. The young produce goose' and FOOT at a similar F2 level and more fronted than the other groups.

[12] Mesthrie (2010b) used the term *backish* to describe a vowel that is produced halfway the dimension central vowel and back vowel, adapted from Watt & Fabricius (2002).

5. All three age groups produce goose' fronter than FOOT (although this is marginal in the young group); thus, FOOT is the backest vowel in this cluster.

The F2 analysis showed that the subset J-words is mainly a high central vowel. FOOT and the subset goose' are high back vowels but with goose' somewhat fronter than FOOT.

5.2.5.2 Vowel quality within GOOSE

Mesthrie (2010b) investigated the front/back dimension of GOOSE in the environment of preceding /j/ as well as preceding coronals and preceding non-coronals. He found that J-words were produced fronter than coronals, and coronals in turn were produced fronter than non-coronals (see also Thomas 2011: 49). Based on Mesthrie's (2010b) study, a second regression analysis was modelled only for F2. This time, goose' was divided into coronals and non-coronals and compared with J-words. The factors entering the model were thus *gender, age group, speech style* and *preceding context* (coronals, non-coronals, /j/ (J-words)) as well as the interactions of *age group*gender*, *preceding context*gender* and *preceding context*age group*. Table 5.58 shows the statistically significant outcome of the model. *Preceding context, gender, speech style* and the interaction of *preceding context*age group* were statistically significant. The interaction of *age group*gender* was not significant, neither was *age group* as a single variable. The R^2c value 0.68 is very high. The fixed effects can explain 42% of the results, and the random effects the remaining 25%. Since this outcome is very similar to that of the F2 analysis for all three variants, the interaction of *preceding context*age group* was not analysed further.

Table 5.58: F2 regression results for goose' (N = 400) and J-words (N = 403) (ph = phoneme, pc = preceding context, ag = age group)

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1590	-0.6330	-0.0415	0.5921	3.9042

Random effects:

Groups	Variance	Std.Dev.
word label	0.1472	0.3837
speaker name	0.0984	0.3136
Residual	0.3060	0.5532

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.6552	0.1764	98.10	-3.713	0.003	**
gender male	-0.3737	0.1206	47.20	-3.100	0.003	**
pc /j/	0.9352	0.1717	81.20	5.448	5.34e-07	***
pc non-coronal	-0.5466	0.2189	111.50	-2.497	0.014	*
style WL	-0.3258	0.0816	581.60	-3.992	7.39e-05	***
pc /j/:ag older	-0.4348	0.1354	709.90	-3.210	0.001	**

R^2m (fixed effects) = 0.42
R^2c (fixed + random effects) = 0.68

```
Post-hoc Tukey, confidence level 0.95:
                       Estimate   Std. Error   z value   Pr(>|z|)
 /j/ - cor == 0          0.9352     0.1717       5.448    <0.001 ***
 non-cor - cor == 0     -0.5466     0.2189      -2.497     0.033 *
 non-cor - /j/ == 0     -1.4818     0.1862      -7.959    <0.001 ***

 WL - IN == 0           -0.3257     0.0816      -3.992    <0.001 ***
 WL - ReP == 0          -0.2330     0.1007      -2.314     0.053

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

The general result is in line with that of Mesthrie's (2010b) (see Section 2.5.2): GOOSE fronting, a once uncommon phenomenon in the English variety of black South Africans, is detectible across the genders and across the age groups. Mesthrie used a different normalisation method than the present study, so a direct comparison with his results is not possible, but the pattern is the same as Figure 5.25 confirms: J-words were produced fronter than coronals, and coronals were produced fronter than non-coronals. It is also obvious that the F2 spread is large, particularly in J-words.

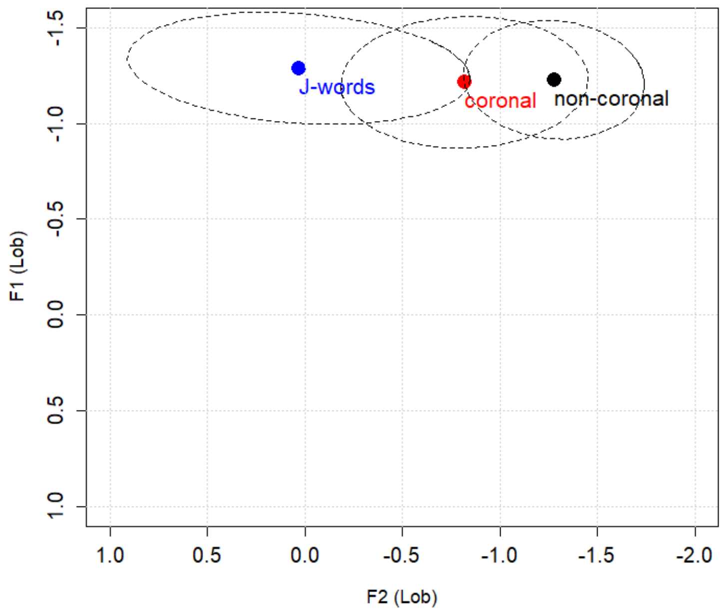

Figure 5.25: GOOSE **by phonological context: J-words (N = 403), coronals (N = 270), non-coronals (N = 130)**

Figure 5.26 shows the outcome in the genders. Females produced fronter vowels in all three environments than males. Mesthrie reports the same for his Black subgroup and concluded that "[b]lack females show the greatest acculturation to the White norm for GOOSE" (Mesthrie 2010b: 28). In his White subgroup, both genders showed GOOSE-fronting, but it was the males who had fronter vowels than the females.

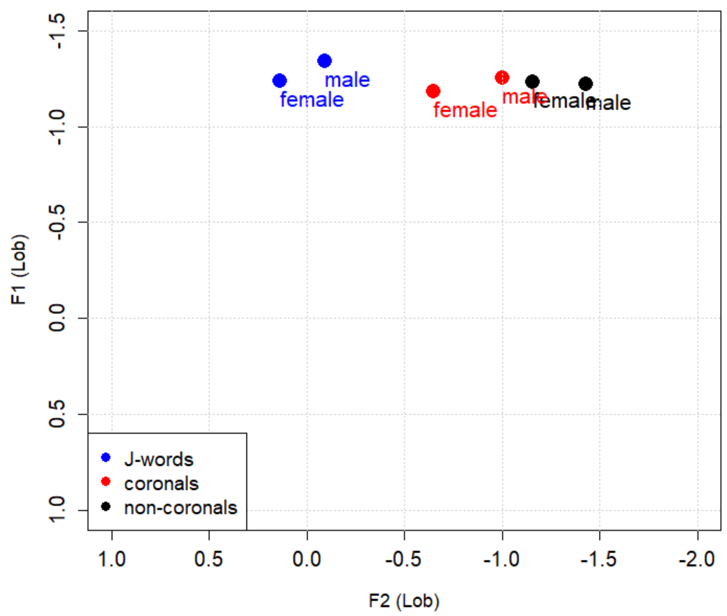

Figure 5.26: The phonological context of GOOSE by gender

Figure 5.27 shows the outcome in the age groups. As can be seen, all age groups differentiate between the three environments. However, the young group shows a higher degree of frontness in all three variants than the other age groups. Further analyses elicited that with one exception, every single speaker irrespective of their age group produced J-words fronter than goose' in coronal and non-coronal contexts (SYF05 had fronter coronals than J-words). And again, with one exception, they also produced coronals fronter than goose' in non-coronal contexts (XYF03 had fronter non-coronals than coronals).

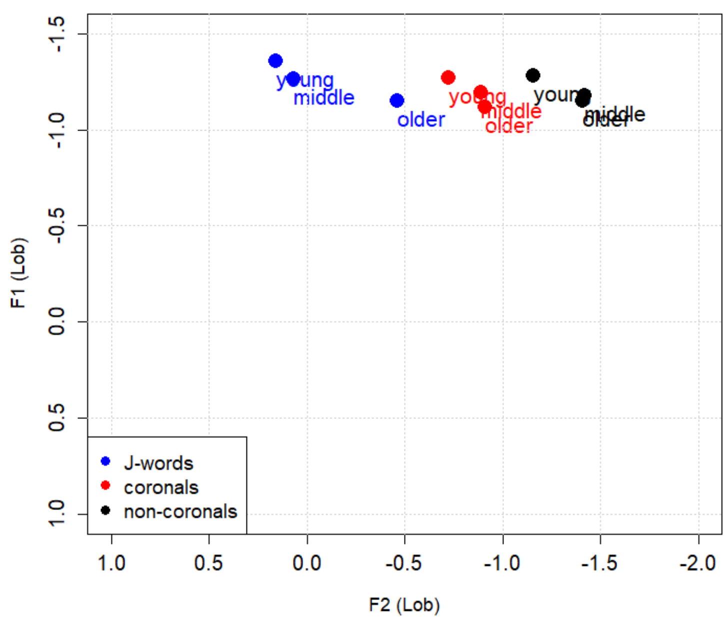

Figure 5.27: The phonological context of GOOSE by age group

Mesthrie also describes that of his twelve black speakers, eight had front values with J-words, three had frontish values and one a central value (Mesthrie 2010b: 17). Mesthrie used a scale suggested by Watt & Fabricius (2002). A proposal for a very rough Lobanov-adapted scale is shown in Figure 5.28:

```
              1.2          0.6          0.0          -0.6         -1.2
&lt;------------ | ------------ | ------------ | ------------ | ------------ | ------------ &gt;
      front      frontish          central            backish      back
```

Figure 5.28: Scale of GOOSE fronting (adapted from Watt & Fabricius 2002)

Taking this scale as a measure, none of the speakers produced front vowels. Four speakers (TYF04, XYF03, ZYM09 and ZMF12) produced frontish vowels, four speakers (ZYM06, XMM09, SOF02 and SOM02) backish vowels, and the remaining thirty-six realised central vowels (see Table 5.59). The results in the present study are not as drastic as Mesthrie's. He examined middle-class university students who had attended private multiracial schools. Some of the families moved to predominantly white suburbs. These participants were exposed to native speaker English very early in life and also started speaking English in their early childhood. At the time of sampling, they were students at the University of Cape Town, a multiracial institution with a high proportion of white students. The lingua franca on campus is certainly English. The chance that the black participants in Mesthrie's sample adopted language features of WSAE seems therefore higher than for the young participants of the present study. They mostly went to township schools, which were entirely black schools, and studied at the North-West University where mostly black students were matriculated.

Table 5.59: Distribution of values for J-words by age group

Age group/Values	front	frontish	central	backish	back
young	0	3	16	1	0
middle	0	1	15	1	0
older	0	0	5	2	0
Total	**0**	**4**	**36**	**4**	**0**

Figure 5.29 shows the outcome according to speech style. Mesthrie (2010b: 18) reports GOOSE-fronting particularly in wordlist style. This is different in the present data: J-words in

WL style are produced fronter than coronals and non-coronals in all three speech styles. However, they are produced much backer than J-words in IN and ReP style. A possible explanation is that speakers took more care of their pronunciation while reading individual words. They may have tried to approximate or unconsciously formed a normative rounded back vowel.

In sum, the fronting of GOOSE in the context of pre-vocalic /j/ (J-words) has become a feature of educated speakers irrespective of their gender and age. However, there is a difference in the degree of fronting between the age groups: Young speakers have the frontest J-words followed by the middle-aged group. The older speakers do not front J-words as strong as the other two groups.

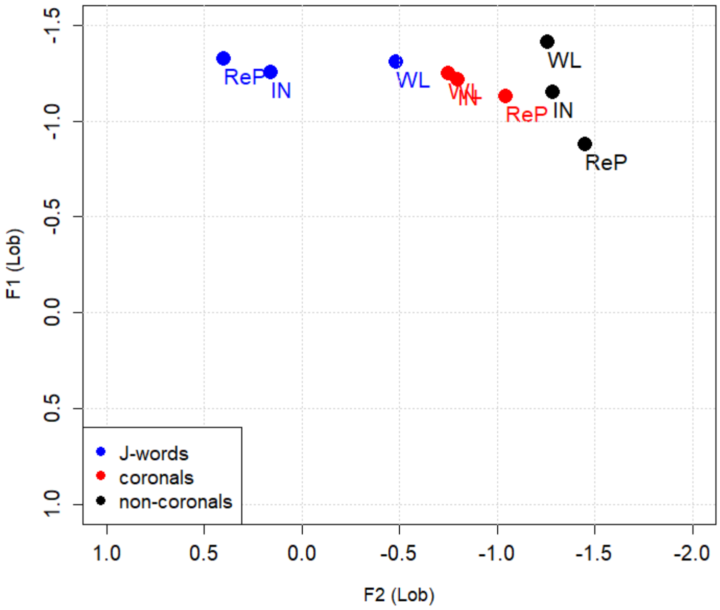

Figure 5.29: The phonological context of GOOSE by speech style

5.2.5.3 Vowel duration

The boxplots in Figure 5.30 show that GOOSE (word-final tokens are excluded) is produced longer than FOOT; the median of GOOSE is higher and the box larger than that of FOOT. What is more, the upper quartile of the GOOSE plot ends at about 140 ms whereas that of FOOT ends at about 60 ms. The lower quartiles end at about the same height. That means that the tokens with a short vowel duration were produced similarly short. Table 5.60 shows that GOOSE produced a very large SD compared to FOOT.

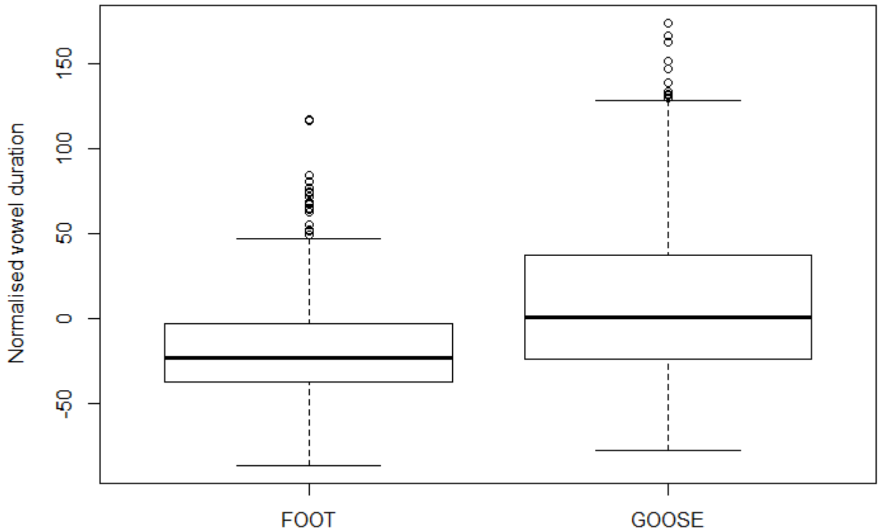

Figure 5.30: Normalised vowel duration (ms) of the high back cluster

Table 5.60: Token numbers, normalised vowel durations and SDs of the high back cluster

Phoneme	N	Mean (ms)	SD
FOOT	356	-12.36	33.94
GOOSE	669	12.12	45.85

The fixed factors entering the model were *phoneme, speech style, gender, age group, following voicing* (voiced, voiceless, pause/gap) and *following manner* (fricative, affricate, plosive, pause/gap) and the interactions of *age group*gender, phoneme*gender, phoneme*age group* and *phoneme*speech style*. The influence of *preceding gesture* (coronals, non-coronals and J-words) were not analysed in the context of vowel duration. The variables that yielded statistically significant results are shown in Table 5.61. These are *speech style* and the interaction of *phoneme*speech style*. Interestingly, *age group* or the interaction of *phoneme*age group* were not statistically significant ($p > 0.1$ and $p > 0.5$ respectively), which means that the groups produced the phonemes similarly long. The differences between FOOT and GOOSE were 23.35 ms in the young, 25.78 ms in the middle and 25.81 ms in the older group. The same applies to gender: The normalised duration in female speakers was -13.36 ms for FOOT and 11.96 ms for GOOSE (25.32 ms difference). The values for the male speakers were -11.01 ms for FOOT and 12.31 ms for GOOSE (23.32 ms difference). The R^2c value of 0.38 is moderately high. The fixed factors explain 16% of the results and the random effects 22%.

Table 5.61: Regression results of duration for FOOT (N = 356) and GOOSE (N = 669) (ph = phoneme)

```
Scaled residuals:
                  Min      1Q      Median     3Q      Max
                -2.7284  -0.6095  -0.1430   0.4540   3.6548

Random effects:
 Groups           Variance   Std.Dev.
 word label        354.470   18.827
 speaker name       12.640    3.555
 Residual         1041.840   32.278

Fixed effects:
                  Estimate  Std. Error   df    t value   Pr(>|t|)
 (Intercept)       -11.742    5.077     96.10   -2.313     0.023 *
 style WL          -16.245    8.421    526.90   -1.929     0.054 .
 ph goose:style WL  45.677    9.751    497.30    4.684  3.63e-06 ***


R²m (fixed effects) = 0.16
R²c (fixed + random effects) = 0.38

Significance codes: p < 0.001 '***', p < 0.01 '**', p ≤ 0.05 '*'
```

The results of the interaction of *phoneme*speech style* are illustrated in Table 5.62. The figure highlighted in grey are the differences within the speech styles. WL tokens of GOOSE are produced longer than all other constellations. Any other interaction did not show significant outcomes. Hundleby (1963) reported that in terms of vowel length, FOOT and GOOSE do not show differences. This is not reflected in the boxplots for FOOT and GOOSE, which are visibly different. However, the statistical model produced only statistically significant results for speech style. It seems that the speakers drew attention to the pronunciation of a prescriptively long vowel. Since there is no difference in gender and age group, this result applies to all speakers.

Table 5.62: Pairwise comparison of duration (ms) for the interaction of phoneme*speech style for FOOT and GOOSE

Interaction	**FOOT*IN**	**FOOT*ReP**	**FOOT*WL**	**GOOSE*IN**	**GOOSE*ReP**
FOOT*ReP	-				
FOOT*WL	-	-			
GOOSE*IN	-	-	-		
GOOSE*ReP	-	-	-	-	
GOOSE*WL	37.01	40.24	54.07	29.19	31.92

Concerning FOOT, Hundleby (1963) reports for Xhosa-English that words spelt with <oo> may be pronounced longer and tenser that those spelt otherwise. This is also true for the present data: Taken at face value, <oo> words are about 20 ms longer than <ou> words and 33 ms longer than <u> words. However, token number and word number per subset may have played a role in this outcome: Of the <oo> words (N = 307), *good* has 108 tokens, *foot* 71 and *cook* (including inflections) 44, comprising 73% of the token number of this subset. Of the <ou, u> words (N = 49), *put* (including inflections) has 37 tokens, which covered 76% of this subset. Apart from that, both subsets are very different in their size (307 tokens to 49 tokens), and they do not differ in their vowel quality. The regression analysis for vowel duration did not yield statistically significant results for the variable *spelling* ($p > 0.08$).

Summing up, it can be said that of the social factors, neither age group nor gender had an impact on the duration of tense and lax vowels. The genders and age groups differentiated between the duration of FOOT and GOOSE, but there were no differences between them. Speech style was the only decisive variable: Speakers reading out individual words (WL) pronounced GOOSE vowels longest. Neither following voicing nor following manner influenced vowel duration. This can be due to the fact that BSAE is a second language where features reported for L1 varieties may not entirely apply. The duration of FOOT was influenced by spelling although not statistically significant. It can be concluded that this feature is apparently very stable although the caveat mentioned above should be considered.

5.2.5.4 Summary of key findings

Regarding vowel quality in the comparison of FOOT and GOOSE as well as in the comparison of the GOOSE variants, the variables *phoneme, gender, age group, speech style* and *preceding context* as well as the interactions of *phoneme**gender, phoneme*age group* and *preceding gesture*age group* showed significant results. GOOSE is produced higher and fronter than FOOT by all genders and age groups. In respect of the vowel quality within the GOOSE set, it can be said that GOOSE fronting is performed in all genders and age groups. This applies particularly to the GOOSE variant J-words, which is produced most fronted, followed by tokens with preceding coronals. All age groups differentiated between J-words, coronals and non-coronals, but the young group showed the highest degree of frontness in all three variants. Concerning gender, females produced fronter vowels in all three environments than males. Regarding speech style, J-words in WL style are produced fronter than coronals and non-coronals in all three speech styles. However, they are produced much backer than J-words in IN and ReP style. The quality of GOOSE is still a back vowel /u/ although with much variation in the front/back dimension. This applies for all three age groups. The young group also performed a classical FOOT vowel /ʊ/. This is also true for the other age groups, but not as systematically as among the young speakers. Therefore, for the middle and older speakers, an emerging /ʊ/ is suggested.

With regard to vowel duration, only *speech style* and the interaction of *phoneme*speech style* showed statistical significances. Interestingly, social factors did not play a role. The speakers differentiated between the duration of FOOT and GOOSE, but there were neither a difference between the genders nor between the age groups. Only speech style showed an impact in that GOOSE in WL style was produced longest. The vowel duration of FOOT seems to be influenced by spelling, but without statistical significance.

5.2.6 "Traditional" and "crossover" speakers

In all vowel clusters, the social variable *age group* showed statistically significant results in both vowel quality and vowel quantity. It was the young group that differed from the middle and older age groups and thus was responsible for this outcome. However, the young participants did not form a homogeneous entity but comprised speakers with quite different performances. This becomes obvious when looking at the Bhattacharyya charts in the Sections 5.2.1.1 to 5.2.5.1. There were six speakers who always or almost always clearly differentiated between tense and lax vowels. These were two men, TYM07 and ZYM09, and four women, SYF05, SYF10, TYF04 and XYF03. These speakers have in common that they either went to a multiracial creche and/or school where the medium of instruction was English, or they grew up in a predominantly white neighbourhood. With this social background, these participants are similar to Mesthrie's (2009b) middle-class BSAE group in his investigation of the GOOSE vowel.

As described in Section 2.5.2, Mesthrie et al. (2015) examined the BATH vowel across four ethnicities and five South African cities. They divided their BSAE sample into four subsamples, two of which they labelled "traditional" and "crossover" speakers (Mesthrie et al. 2015: 22). Adopting this terminology, the label "crossover" is valid for the six young speakers describes above. The label "traditional" can be applied to the remaining 38 speakers, partly in terms of school education and partly because of their vowel performance. A vowel plot of the "crossover" speakers in Figure 5.31 reveals a higher number of contrastive monophthongs than the vowel plot for the whole cohort (see Figure 5.1). Genuine vowel clusters can only be discerned in three subsets: The first comprises the tense vowels of the back cluster THOUGHT, NORTH and

FORCE located to the right of FOOT. They are produced as one phoneme, also reflected in the practically identical standard deviations. Their quality is a close-mid back vowel. The somewhat confusing cluster of back vowels below consists of two different clusters. The first includes the lax back vowels LOT and CLOTH. Their means are the two dots top left of the cluster. Their quality is an open rounded back vowel. The second cluster contains the tense back vowels BATH, START and PALM. Their quality is a raised open unrounded back vowel. The means of BATH and START in the middle of the cluster lie close together, and the overlap of their SDs is high suggesting one phoneme. PALM (including the <lm> subset) is produced somewhat backer. Backing and raising of BATH as described for WSAE and the crossover speakers in Mesthrie et al.'s (2015) study is also a feature of the crossover speakers in this work.

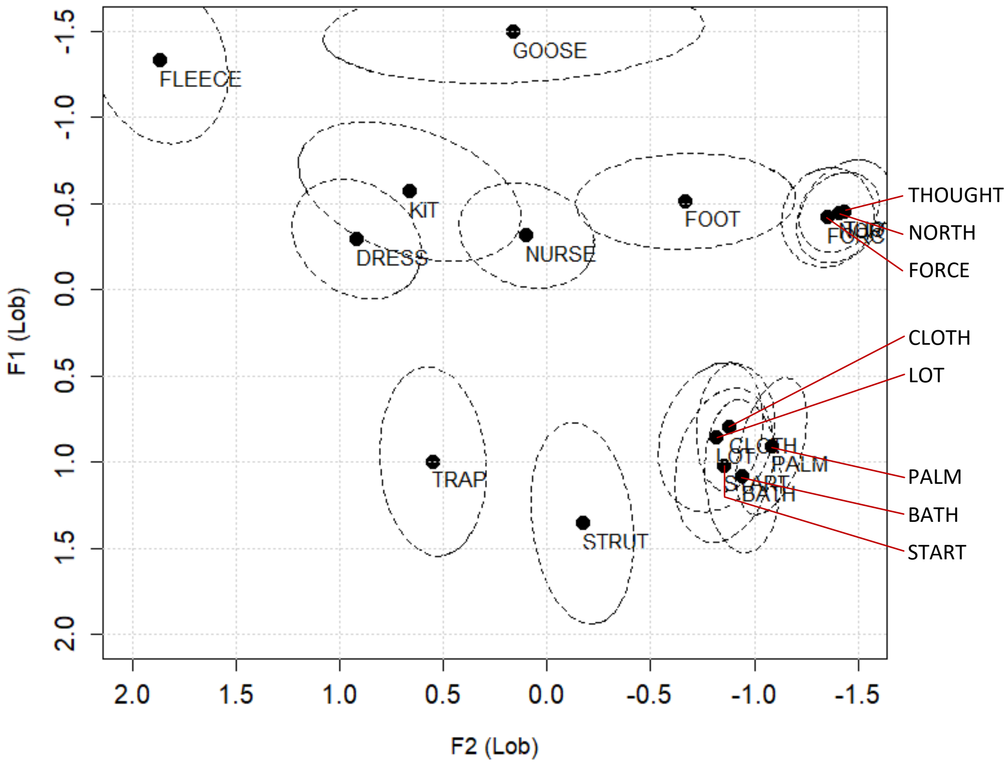

Figure 5.31: Lobanov-normalised vowel space of BSAE across the "crossover" speakers with 1 SD (N = 1,165)

In this vowel plot, the quality of the phonemes KIT and GOOSE are particularly interesting. KIT includes centralised [ï] and close-mid central [ɘ] tokens. The "virtual wholesale centralization of KIT" as reported for young middle-class female WSAE speakers by Bekker (2008: 277) can be observed in young black crossover speakers as well. GOOSE shows a high degree of fronting and is mainly produced as a central vowel. In both cases, the prestige white language model was adopted by these speakers. Based on the above vowel plot, Table 5.63 provides the vowel inventory of crossover speakers and suggests eleven vowels: /i, ɪ, e, ɘ, æ, ɐ, ɑ, ɒ, o, u, ʊ/. Since the vowel contrast is very high, this system is referred to as acrolectal BSAE. Van Rooy (2004) suggested a system of nine vowels with two emerging phonemes. The present results mainly correspond with his findings although the actual quality of single phonemes differs.

Table 5.63: The stressed monophthongs of acrolectal BSAE

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i]	PALM	[ɑ]
KIT	[ɪ̈], [ɘ]	LOT	[ɒ]
DRESS	[e]	CLOTH	[ɒ]
NURSE	[ɘ]	THOUGHT	[o]
TRAP	[æ]	NORTH	[o]
STRUT	[ɐ]	FORCE	[o]
BATH	[ɑ]	GOOSE	[y], [ʉ], [u]
START	[ɑ]	FOOT	[ʊ]

The “traditional” speakers were plotted separately as presented in Figure 5.32. In contrast to the crossover speakers, this vowel plot very much resembles that of the whole population in Figure 5.1. There are five vowel clusters: high front, mid front, low central, mid back and high back. The lexical set PALM also contains the <lm> subset. Table 5.64 provides the vowel inventory of the traditional speakers. Although there is more allophonic variation, the vowel contrast is much lower than that of the crossover speakers. For the traditional speakers, a five-vowel system is suggested: /i, ɛ, ɑ, ɔ, u/ with one emerging phoneme /ʊ/. Van Rooy (2004) also defined five vowels for his sample of mesolectal speakers, for which reason this vowel system is referred to as mesolectal BSAE.

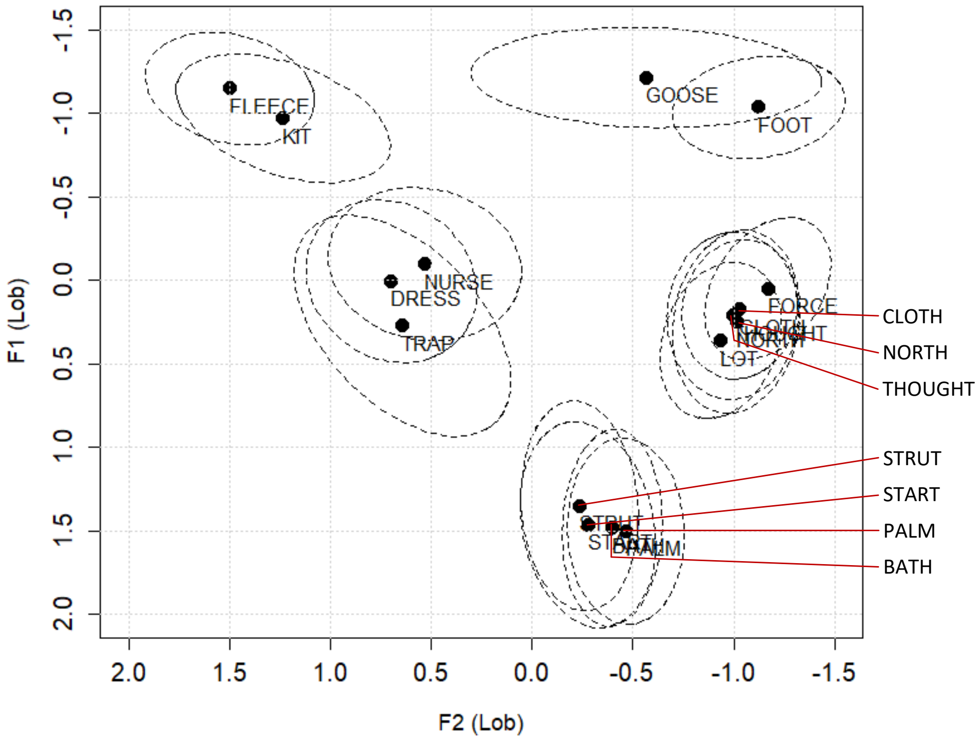

Figure 5.32: Lobanov-normalised vowel space of BSAE across the "traditional" speakers with 1 SD (N = 7,215)

Table 5.64: The stressed monophthongs of mesolectal BSAE

Lexical set	Allophones	Lexical set	Allophones
FLEECE	[i] > [ɪ]	PALM	[ä], [ɐ]
KIT	[i], [ɪ] > [ɪ̈]	LOT	[ɔ]
DRESS	[e], [ɛ]	CLOTH	[ɔ]
NURSE	[e], [ɛ], [ɘ]	THOUGHT	[ɔ]
TRAP	[æ], [ɛ] > [a]	NORTH	[ɔ]
STRUT	[ä], [ɐ]	FORCE	[ɔ], [o]
BATH	[ä], [ɐ]	GOOSE	[ʉ], [u]
START	[ä], [ɐ]	FOOT	[u], [ʊ]

The speakers investigated in this study were finally divided into crossover speakers and traditional speakers with regard to their vowel inventory. The captions of the two tables above, however, refer to acrolectal and mesolectal pronunciation to enable a direct comparison with Van Rooy's (2004) classification (see Section 2.5.1). As mentioned before, acrolect, mesolect and basilect are regions on a second language continuum. Schmied (1991: 47), who proposed these labels for the discussion of African English varieties, defines education (in particular the length and degree of formal education in English) and occupation (that is, the necessity for and amount of English used in everyday life) as the two key factors for a person's level of English proficiency. A typical acrolectal speaker has received 14 years of English education, has studied at a university and works for example as a university lecturer, a medical doctor or a senior officer in the civil service (Schmied 1991: 48). With regard to the speakers investigated, these conditions apply to ten middle-aged and two older participants. Formally, they could thus be allocated to the acrolect, but in terms of their vowel performance, these speakers were definitely mesolectal. All participants in the present study had passed at least the secondary school examination and were fluent speakers of English, but a high English proficiency is not necessarily linked to a native-like vowel inventory. Only the six crossover speakers who were schooled with English as the medium of instruction or grew up in a white neighbourhood made use of a vowel repertoire similar to that of a native speaker. They can be unambiguously regarded as acrolectal. It seems that an early exposure to mother tongue English, in both education as well as private conversation, is the key determinant of a pronunciation that is close to the respective standard English variety. For the present sample, occupation was not a cue in this respect. Currently, the occupation of a speaker probably not always co-determines his or her English language skills, but with the growing number of young well-educated Blacks who may take up jobs in high positions, the correlation between occupation and English proficiency will be more established in the future.

5.3 Validation of the research hypotheses

In consideration of the above results, Table 5.65 summarises the research hypotheses and the decisions. The hypotheses H1.2, H2, H2.1 and H3 could be straightforwardly rejected since the comparison of the variants of the respective variable did not yield statistically significant differences. The remaining hypotheses were more difficult to validate because the outcome was not always unambiguous.

Table 5.65: Validation of the research hypotheses

Hypothesis		Validity
H1	The phonologies of the age groups are significantly different.	**Partly accepted**. See hypotheses H1.1 to H1.3.
H1.1	The youngest age group differ significantly from the other two.	**Accepted** only when comparing the age groups as entities. Of the twenty young

Hypothesis		Validity
		speakers, only six had a broader vowel inventory. They were responsible for this difference.
H1.2	The middle and older age group differ from each other but to a lower extent than to the youngest age group.	**Rejected** because the middle and older age group mainly produced the same results.
H1.3	The younger the age group, the higher the number of distinct vowels.	**Accepted** only when comparing the age groups as entities.
H2	Gender differences within the age groups are significantly different.	**Rejected** because males and females mainly produced the same results.
H2.1	Young females differ significantly from the other participants.	**Rejected** because the interaction of *age group*gender* yielded statistically significant results only for the duration of FLEECE and KIT, but young females were not involved in this outcome.
H3	Speakers whose native language has a seven-vowel system have more distinct vowels in English than those who have a five-vowel system.	**Rejected** because the different vowel systems of the substrate languages did not show significant differences whatsoever.
H4	The vowels in the speech styles show significant differences.	**Partly accepted**. Most WL tokens differed from both ReP and IN tokens while ReP and IN did not markedly differ from each other.

The aim of the present work was the description of the BSAE monophthong vowel system. It investigated the linguistic and social factors that influenced pronunciation. Many results came out as expected, such as the influence of speech style on pronunciation or the lacking influence of the spelling variants of NURSE. But some findings were different from or at least not completely congruent with what is suggested in the literature. In this respect, two results are particularly interesting: the performance of the KIT split and GOOSE-fronting. These middle-class WSAE features were found in males and females and, most interestingly, in all age groups although to different degrees. Since this work is the first that also considered middle-aged and older speakers and thus cannot refer to previous observations, it is difficult to say how old these features actually are. Considering that adults do not easily change their pronunciation, it appears that the two phenomena have been present for more than one generation. In this respect, the question must be raised whether these features are truly "White" or just a result of a universal development.

The research method of this work was an apparent time study as a means for the detection of language change in progress. Whether language change is on the way indeed can neither be straightforwardly confirmed nor rejected. According to the statistical analysis, change is going on, and it is the young that differs irrespective of their gender. Zooming into this group, however, of the twenty young participants, only six crossover speakers were the tip of the scale for this result. The majority of the sample, including 14 young speakers, exhibited a very similar, traditional pronunciation. Apart from that, concerning the number of vowels, both subsamples are comparable to Van Rooy's (2004) summary of the acro- and mesolectal BSAE vowel space. Hence, on the one hand, it seems safe to say that with the participants of this study, language change could not be observed. It can therefore be concluded that in the mesolect, where all three age groups were represented, the phonemes are stable in apparent time. It can also be concluded that educated BSAE is still majorly mesolectal.

On the other hand, concerning the particular vowel qualities, there are hints for a change in BSAE, and this change is obviously going on not only in South Africa's big cities, as described for example by Da Silva (2007), Mesthrie (2009b) and Mesthrie et al. (2015), but also in smaller urban areas. Again, KIT and GOOSE shall be cases in point: GOOSE fronting is in operation and may be numerically more salient in the future, particularly in the younger generation. As for KIT, the split may remain stable, but it is also possible that KIT will take part in a reverse shift as reported by Chevalier (2019) for WSAE.

Tackling the question of the influence of a de jure unsegregated society on BSAE pronunciation, it is certain that the language change described above could only come about in this very atmosphere. However, as the majority of the young speakers proved, not everyone has had the opportunity to use "integrated educational facilities" (Bekker & Van Rooy 2015: 296), i.e. to attend multiracial schools, or to be instructed by native speakers of English. In this context, another aspect could have played a part in the outcome of the study: Speakers of BSAE are numerically dominant. This dominance can form the basis for a gradual acceptance of local forms of English, which, in turn, is an expression of a "locally rooted linguistic selfconfidence“ (Schneider 2007: 49). Since multilingualism is the norm in South Africa, BSAE speakers can choose from a number of languages and may not necessarily go for English as an identity marker (Schneider 2007: 175). This mindset can also only prosper in a liberal socio-political climate. Hence, there are signs that the present political system in South Africa has an impact on the development of BSAE indeed.

5.4 Further observations

The focus of the study was on prescriptively stressed monophthongs in controlled phonological environments, but the intense listening to the sound files revealed some interesting auditory observations along the way, which were not included in the acoustic analysis. These are impressionistic, remain largely uncommented and have no claim to completeness. Some incidents are anecdotal, but some seem to have occurred more systematically.

One feature was vowel length. The Reading Passage (ReP) contained the minimal pair *shot* and *short*, it means that both words differ in one phoneme, the lax [ɒ] in *shot* and the tense [ɔː] in *short*. Interestingly, 37 speakers (including five crossover speakers) produced *short* shorter than *shot*, and 19 speakers produced *short* shorter than the shortest duration of *shot*, which was 93.0 ms. Maybe, it was the word *short* itself that triggered a short pronunciation, indicating that the lexical meaning of a word may be stored in the brain along with the pronunciation of this word. Another observation concerning vowel length, in combination with vowel quality, was the ReP word *fields*. Twenty-one speakers pronounced *fields* as [fiːldz], with a long vowel. Thirteen speakers, however, produced it with a short vowel or syllabic /l/, such as [fɪldz], [fɘldz] or [fl̩dz]. In contrast to the frequently observed absence of lax vowels, it seems that post-vocalic /l/ shortens and centralises the long /iː/. One speaker read [feld], which was certainly influenced by *veld* [feld], the Afrikaans word for *field*.

With respect to sonorants, the pronunciation of /l/ and /r/ showed some variation. In words like *exactly, fluently, clean, later* or *village*, /l/ was often produced as the lateral voiceless fricative [ɬ] and the affricate [tɬ]. Both are features of Bantu languages (f.e. Arellano 2001: 10) and will have provoked this pronunciation. Rhotic /r/ was often produced as a trill [r] probably borrowed from the substrate languages. Young speakers also pronounced an approximant [ɹ]. This finding conforms with Van Rooy’s (2004) outcome. Hartmann & Zerbian (2009) also observed a high number of [ɹ] realisations in coda position, particularly among young affluent females, but they found very few trills.

The simplification of consonant clusters, in particular, the deletion of plosives was another observation. Examples are *act* [æk], *exactly* [ɪgzæklɪ], *accept* [asep], *don't bother* [dɔn bɔðer], *escaped from* [eskeɪp frɔm]. The deletion of plosives is also reported by Van Rooy (2007: 32). The affricates /tʃ/ showed variation and was often produced as [ʃ] in words like *much, watch, change* and *village* (see also Van Rooy 2004: 950).

Epenthetic schwa, as observed by Simo Bobda (2000a) in several African English varieties, also occurred in the present data, in particular in ReP: "*… a dark forest[ə]near the foot of a mountain*", "*… even stayed[ə]with him …*".

Stress shift also occurred frequently. In ReP, the words *company, concern, overcoming* and *neighbourhood* were often pronounced *comˈpany, ˈconcern, overˈcoming and ˌneighbourˈhood.* The stress shift is a transfer from the Bantu substrates where penultimate syllables are lengthened as is the case with the first three words (Hundleby 1963). If the final syllable contains a tense vowel and a coda consonant, as in *neighbourhood*, the stress is placed on the final syllable (Hundleby 1963; Lanham 1990; Van Rooy 2000, 2004).

Finally, postalveolar clicks [!], another feature of the native languages, were frequently performed during the interviews and sometimes while reading. They were produced mainly but not exclusively by older speakers.

6 Conclusion

6.1 Overview of the project

The core of this work is a sociophonetic examination of contemporary Black South African English, BSAE, the most spoken second language variety in South Africa. It investigates the vowel quality and vowel duration of prescriptively stressed monophthongs in a socially stratified sample of 44 BSAE speakers. Data collection took place in the South African provinces Gauteng and Free State. The sample consisted of 21 females and 23 males. The participants were divided into three age groups named young (19–23 years old, N = 20), middle (25–54 years old, N = 17) and older (58–84 years old, N = 7). The participants were audio-recorded in three speech styles with different degrees of formality: interview style (IN), reading style (ReP) and wordlist/citation style (WL). This setup was employed to determine the number of distinct vowels of BSAE, to determine durational differences between prescriptively tense and lax vowels and to investigate the possible influence of linguistic and social factors on the variation in BSAE.

Around 14,000 tokens of Lobanov-normalised formant values were extracted and subjected to descriptive and analytical statistics. Each vowel cluster was analysed for the frequency of the formants F1 and F2 and for vowel duration. Vowel overlap was calculated with the Bhattacharyya coefficient (BC) and the Pillai score. Vowel length was compared by normalised vowel duration. The regression analysis was carried out with linear mixed effects models. Random factors in each regression run were *speaker* and *word label*. The fixed factors, i.e. the variables examined, were *phoneme, phonological context, spelling, speech style, vowels in L1 language family, gender* and *age group*. The fixed factors can be divided into linguistic factors (*phoneme, phonological context, spelling, speech style* and *number of vowels in L1*) and social factors (*gender* and *age group*).

Regarding vowel quality, the fixed factors were *phoneme* (selection according to the vowel cluster), *spelling* (for the analysis of NURSE, STRUT and FOOT), *speech style* (IN, ReP, WL), *number of vowels in L1* (5, 7), *gender* (male, female), *age group* (young, middle, older) and *phonological context*. The factor *phonological context* was divided into *prevocalic context* and *adjacent context*. The *prevocalic context* included the variants coronals and non-coronals. For the analysis of GOOSE, the variant J-words was added. The analysis of KIT contained the variable *adjacent context* with the variants high front and centralised. The interactions of factors comprised *age* *group*gender, phoneme*gender* and *phoneme*age group*, and for the analysis of GOOSE additionally *preceding context*age group* and *preceding context*gender*.

Regarding vowel duration, the fixed factors were those above. Again, of the preceding context, the variants coronals and non-coronals were always included as well as J-words for the analysis of GOOSE. Apart from that, *following voicing* (voiced, voiceless, pause/gap) and *following manner* (fricative, plosive, affricate, pause/gap) were added. The variable *number of vowels in L1* was excluded. The interactions above were added by *phoneme*speech style*.

6.2 Summary of the results

In essence, across all variables, the results show a general presence of the tense/lax distinction. The social variable *age group* had the biggest influence on both vowel quality and quantity. The young age group differed significantly from the middle and older group. The young speakers showed the greatest variation ranging from complete overlap of two vowels to clear vowel distinction. The social variable *gender* rarely yielded significant results. The KIT split, reported for White South African English (WSAE), was performed by speakers of all age groups and genders. Another outcome was GOOSE fronting in the context of preceding /j/, which could also be observed in all age groups and genders. Of the linguistic variables, *speech style* was the most decisive, showing that most participants differentiated between formal style (WL) on the one hand and less formal (ReP) and casual style (IN) on the other. The *number of vowels in L1* showed no significant results whatsoever. The following list of the variables employed describes the findings in more detail.

Phoneme: Phonemes were analysed entirely in interactions. This made sense because they were always uttered by someone and in a certain context. Yet, at this point, it is worth having a brief look at them as single variables. In general, prescriptively tense and lax phonemes within a cluster differed statistically significantly in vowel quality and vowel quantity. There were only two exceptions concerning vowel duration: First, START was produced significantly longer than BATH. Second, the vowel duration of STRUT and BATH did not differ.

Phonological context: Preceding context only played a role in high vowels. In the GOOSE-FOOT cluster, J-words were produced frontest, followed by coronals; non-coronals were pronounced backest (see Mesthrie 2010b). In the FLEECE-KIT cluster, non-coronals were produced fronter than coronals.

The adjacent context determined the pronunciation of KIT: Particular neighbouring environments that trigger the split of KIT into front and centralised variants in WSAE was also present in BSAE. Hence, the KIT split is a feature of BSAE performed by speakers of all age groups and genders.

Following context was decisive for the pronunciation of PALM. Post-vocalic bilabial nasal /m/ triggered a backer pronunciation than other contexts. This outcome was not statistically significant but is nevertheless in line with previous reports (Di Paolo et al. 2011: 88; Thomas 2011: 101). Another reason for the backing of the vowel can be the historically pronounced /l/ in the sequence <lm> (Mesthrie et al. 2015).

Following manner influenced vowel duration in three clusters. In the high front cluster, vowel length increased in the order affricates < plosives < fricatives. In the mid back cluster, vowels before fricatives and plosives were produced longer than before affricates. In the low central cluster, vowels before fricatives and affricates were longer than before plosives. These results are only partly in line with those of Peterson & Lehiste (1960) who observed the longest vowel duration before voiced fricatives, but also found that affricates and plosives affect preceding vowels in the same manner.

The voicing of a postvocalic consonant was only decisive in the mid back cluster where vowels followed by voiced consonants were produced longer than vowels followed by voiceless consonants. This result is consistent with the findings of former research (e.g. Peterson & Lehiste 1960; House & Fairbanks 1953 and Denes 1955 in Thomas 2011), but it is surprising that voicing did not play a role in the other clusters. This may be based on the substrate languages of the speakers. Bantu languages know voiced and voiceless consonants, but not all voiceless consonants that have a voiced counterpart in English exist in the Bantu languages. For example, Tswana has no voiced fricatives, and the only voiced plosive consonant is /b/ (Arellano 2001: 10–12).

The phonological context also influenced the pronunciation of FORCE. In word-final position, FORCE was produced backer than THOUGHT, NORTH and FORCE in non-word-final position.

Spelling: The different spelling variants of NURSE and STRUT did not influence vowel quality and vowel duration respectively. For NURSE, this corresponds with the findings of Simo Bobda (2000b) for various L2 varieties of Southern Africa. The STRUT vowel has not been investigated yet in this respect. A comparison with other findings is hence not possible. The duration of FOOT seems to be influenced by spelling in that <oo> is produced longer than other spelling variants. This result is confirmed by Hundleby (1963), however, the present outcome was not statistically significant.

Speech style: Regarding speech style, it was the WL style that mainly differed from ReP and IN style whereas the vowel quality of ReP and IN tokens was often very similar. WL tokens were produced fronter than ReP and IN tokens in the high front cluster and lower in the mid front, low central and mid back cluster. In terms of vowel length, WL tokens were often produced longer than tokens of the other speech styles although not always statistically significant. The longer WL tokens (no matter whether they included a tense or a lax vowel) may hint at the fact that the speakers took more care when pronouncing single words (as the most formal style) compared to reading a text (ReP) or making conversation (IN). This is probably a typical process that may take place subconsciously rather than intentionally.

Number of vowels in L1: The number of vowels in L1, which refers to the membership in language families with different vowel systems, did not yield significant results at all (cf. Van Rooy & Van Huyssteen 2000).

Gender: Gender-based results were only partly in line with former research. Male and female speakers rarely differed in their realisation of vowels, be it in vowel quality or in vowel duration. Exceptions were the high vowels: Males produced FLEECE and KIT slightly higher and shorter than females. Females showed a greater KIT-FLEECE distance than males and thus a greater vowel distinction. Regarding the KIT split, both speaker categories produced tokens of the centralised variant backer than those of high front contexts; females, however, exhibited a larger distance between the two variants. In the high back vowels, the GOOSE variant of J-words was produced most fronted in both genders, but females produced FOOT and GOOSE fronter than males, that is, they also showed a more advanced GOOSE-fronting than males. This outcome is identical with that of Mesthrie's (2009b) study. It seems that female speakers, at least in parts, pursue prestige norms. By and large, however, men and women performed similarly. Therefore, gender-based expectations may be not universal in second language varieties or only valid for middle-class environments.

Age group: Age group was frequently the decisive factor. To be more specific, it was the young group that mainly differed from the middle and older group whereas comparison between the middle and the older group rarely yielded significant differences. In the high front cluster, FLEECE was produced fronter than KIT, but members of the young group showed the longest distance in vowel height and frontness between FLEECE and KIT. In terms of duration, within and across all age groups, FLEECE was pronounced longer than KIT. Regarding the KIT split, each age group produced two variants (high front and centralised), but the young group had the largest distance between the two variants. In terms of duration, within and across all age groups, FLEECE is pronounced longer than KIT. In the mid front cluster, young speakers produced NURSE significantly backer than the other age groups. They also showed the largest distance between the phonemes, that is, they produced DRESS frontest, NURSE backest and TRAP lowest. NURSE was generally pronounced longer than DRESS and TRAP. In the low central cluster, the young speakers did not differ in their quality of the tense phonemes BATH, START and PALM (except the <lm> variant), but their quality differed from that of STRUT in that STRUT was pronounced fronter and higher. The middle and older age group produced all four phonemes as one. Concerning duration, in all age groups, START was longer than STRUT and BATH, and STRUT and BATH did not differ in their length. The young group showed the biggest difference in vowel length between STRUT and START. In the mid back cluster, the young differentiated between lax and tense vowels, but only in vowel quality. Speakers of the middle and older age groups showed little and no difference respectively and produced just one sound for both phoneme groups. Regarding vowel duration, there was no clear-cut result in any age group. Finally, in the high back cluster, GOOSE is produced fronter than FOOT by all age groups. The GOOSE allophone J-words is a central vowel in the young and middle group. All age groups distinguished J-words, coronals and non-coronals, but the young produced all three variants fronter than the other groups. Young speakers showed the largest difference in vowel height between GOOSE and FOOT. Regarding vowel duration, GOOSE was produced longer than FOOT by all age groups.

A closer look at the young group revealed that it was not the whole group who was responsible for this outcome, but only six speakers labelled “crossover” speakers (Mesthrie et al. 2015). They had in common that they either attended a multiracial creche and/or school with English as the medium of instruction, or they grew up in a predominantly white neighbourhood. These speakers employed an acrolectal inventory of eleven vowels: /i, ɪ, e, ɘ, æ, ɐ, ɑ, ɒ, o, u, ʊ/. This result mainly corresponds with Van Rooy's (2004) findings. The remaining 14 young speakers along with all speakers of the middle and older age group formed the subsample of the “traditional” speakers (Mesthrie et al. 2015). They exhibited a mesolectal five-vowel system: /i, ɛ, ɑ, ɔ, u/ with one emerging phoneme /ʊ/. Again, this result mainly corresponds with that of Van Rooy (2004).

6.3 Limitations

This study has some limitations regarding the sample, the data collection and the data analysis. Concerning the sample, the initial aim to apply the method of stratified sampling had to be discarded. Instead, the participants formed a convenience sample as mentioned in Section 4.1. The disadvantage of this sample was that the number of participants for each age group differed greatly and so did the number of men and women within the age groups. The latter was particularly the case in the older group, which contained four female and only two male participants. A statistical analysis has greater power if an equal number of participants is assigned to each condition (e.g. Robins et al. 2010), i.e. if all age groups contain an equal number of speakers, and each age group contains an equal number of genders. In general terms, every cell of a bi- or multivariate matrix should ideally contain the same number of subjects.

Concerning the data collection, the sociolinguistic interview was sometimes shorter than recommended in the literature. According to Labov (1984b: 32) an interview should take at least 60 minutes to obtain sufficient data and to reduce the phenomenon of the observer's paradox. The interview part of the whole recording session in this study lasted between 25 minutes and 65 minutes. The recordings took place in working environments, which means most of the participants worked or studied there; they saved their lunchtime or a free period to be interviewed. Longer interviews may have yielded more data. However, and although all participants took part in the study voluntarily, their time spent for the interviews was considered more valuable than the duration of the recordings and the amount of speech data. Therefore, it was tried to avoid overburdening the participants with extensive interview sessions. In this context, it should also be mentioned that the interviewer was white and a foreigner to South Africa. In a multicultural society like South Africa's, this should have been only a minor issue, if any, but could still have had an impact on the pronunciation of the participants.

Furthermore, some results were difficult to interpret due to a lack of more detailed information about the sociolinguistic background of the participants. An extended investigation in this respect would have possibly shed more light on the question why the speakers behaved the way they did. For example, questions about the type of nursery school and school the participants attended as well as the language of instruction in these institutions were frequently asked but not comprehensively enquired. The systematic inclusion of such questions in the questionnaire might have been useful to identify sociolinguistic patterns.

A drawback in terms of data analysis refers to the mathematics of the statistical models. The computing tool R and the scripts (multiple linear mixed effects regression (lmer) models) employed for the statistical modelling are considered appropriate for the kind of statistical analy-sis in this work. While it was possible to read and interpret the output of the models, it was impossible to understand the computational decisions that led to this output. Perhaps, it is not necessary to fully comprehend the statistical formulas, but an insight into the calculations 'in the background' would increase the confidence in the results.

These shortcomings should not diminish the relevance of the present work: It is the first acoustic-phonetic analysis of the monophthong system of BSAE using normalised formant data, and it is one of the first studies that compares the performance of BSAE speakers of different age in apparent time. This work therefore contributes to the knowledge of BSAE by reporting new findings and confirming previous ones respectively.

6.4 Outlook

The present work provides an overview of the vowel system of contemporary educated BSAE and the social and linguistic variables responsible for variation within this variety. Yet, it also leaves room for further research:

1. This analysis concentrates on prescriptively stressed monophthongs, but the monophthong inventory of L1 English also includes unstressed vowels represented by the lexical sets commA, lettER, happY, horsES and About (Wells 1982a). If unstressed vowels do not exist in the substrate language, they often remain unreduced and are approximated to the nearest fitting stressed vowel. In the case of schwa /ə/, Van Rooy & Van Huyssteen (2000: 20–21) distinguished at least seven different schwa allophones in BSAE depending on their position in the words, and Mesthrie (2017) found instances of deracialised schwa. Delving into the realm of the unstressed vowels would certainly be a rewarding task.
2. The same applies to the investigation of diphthongs. Previous works reported monopthongisation of diphthongs, in particular in the lexical sets of SQUARE, NEAR and CURE in both the mesolect and acrolect. These diphthongs include a central vowel as their offset, which is mostly avoided in this context (Van Rooy 2004: 946). Whether or not this is still the case, in particular among acrolectal speakers, who make use of central vowels otherwise, has not been investigated recently.
3. The vocalic environment examined was limited to phonological contexts that were comparatively easy to distinguish from the vowels in question and which are reported to have little impact on vowel quality. For this reason, post-vocalic /l/ (as only touched upon in Section 5.4) was excluded, but its effect on the preceding vowel would also be worth investigating further.

4. Research has proven that language change is led by women. It has also proven that women adopt incoming prestige forms faster than men. In the South African context, this has been exemplarily shown with the degree of GOOSE fronting (e.g. Mesthrie 2009b, 2010b). In the present study, however, the interaction of *age group*gender* yielded only one statistically significant result, suggesting that generally, there was no difference between male and female speakers across the age groups. Adding the variable *phoneme* to analyse the interaction of *phoneme*age group*gender* may have a chance to elicit significant results. To accomplish this, a larger data basis is needed though.
5. Sociolinguistic research can rely on more and more sophisticated analytical tools. For example, so-called conditional inference trees (ctrees), also known as decision trees, are frequently used in addition to regression models (e.g. Chevalier 2019), or to predetermine the variables for these models. Ctrees visualise the interplay of variables as a tree with binary branches in a hierarchical order starting with the most influential factor. Statistically not significant factors are not depicted and can therefore be omitted in the statistical models. Figure 6.1 illustrates a ctree of F2 for KIT carried out with the computing software R, package ‘party’ (Hothorn et al. 2020). The random factors are *adjacent context* (high front, centralised), *age group* (young, middle, older) and *gender* (male, female). The most

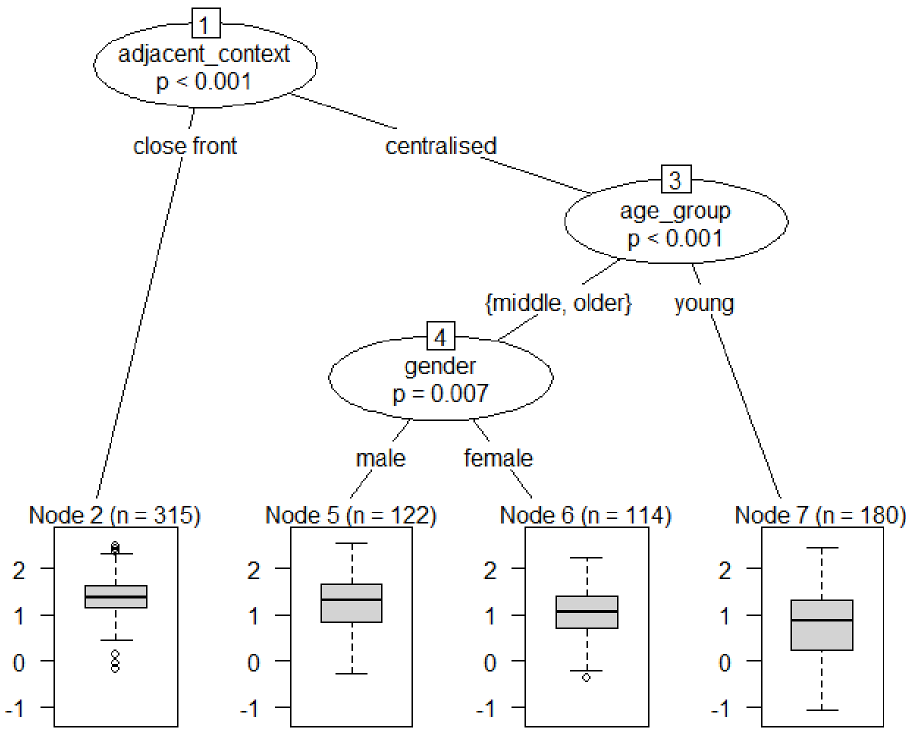

Figure 6.1: Ctree of KIT: F2 for adjacent context, age group and gender

influential variable is *adjacent context*. The high front variant is not divided further indicating that all speakers irrespective of their age group or gender produce a very similar high front KIT variant. Concerning the centralised variant, age group is relevant and splits into middle and older speakers, and young speakers. In the middle and older group, gender plays a role whereas in the young group, males and females do not differ in their production of centralised KIT. The boxplots at the bottom show the realisation of KIT for the final nodes. The scale shows the Lobanov Units. The lower the boxplots, the backer the vowels. This useful tool was not employed in the present data analysis, but in future studies, it can help to select the variables for further analysis.

6. A final note shall refer to the technical aspect of speech analysis. New software technologies for vowel formant measurements have improved the efficiency and accuracy of the annotation and segmentation of speech. Meer et al. (under review) compared several automatic methods for vowel formant prediction, among others two FAVE methods (Forced Alignment and Vowel Extraction) and formant ceiling optimisation methods on the basis of Trinidadian English. They concluded that these methods provide high-quality measurements and even outperform popular speech analysis software, such as Praat. Therefore, the authors recommend them for the phonetic analysis of New Englishes. Research on vowel variation in BSAE would thus also profit from these new technological methods.

The introspection of BSAE should be continued, but also its comparison with other South African varieties. Most research in this respect has been carried out in larger cities because they provide a favourable environment for language change, but fieldwork should be expanded to smaller towns or rural areas. What is more, since the South African society is becoming increasingly diverse, the focus should be on both regional and social dialectology. Mesthrie et al. (2015) started in this direction with the examination of the BATH vowel and found incipient regional differences in BSAE. The study of other vowels in this manner should follow. As Mesthrie et al. (2015: 27) note: “The sociophonetics of South African English is a wide-open field” to which I can only reply: “Eish, that’s true.”

References

Adank, Patti, Roel Smits & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116(5). 3099–3107.

Arellano, Jorge. 2001. *The sound system of Setswana.* Gaborone: Lightbooks.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Bamgboṣe, Ayọ. 2009. A Recurring Decimal: English in Language Policy and Planning. In Braj B. Kachru, Yamuna Kachru & Cecil L. Nelson (eds.), *The handbook of world Englishes*, 645–660. Malden, Mass: Wiley-Blackwell.

Bantu Zone S. 2014. *MultiTree: A digital library of language relationships.* Bloomington, IN: Department of Linguistics, The LINGUIST List, Indiana University. http://multitree.org/codes/cnbs.

Barnard, Etienne & Daan Wissing. 2008. Vowel variation in Southern Sotho: an acoustic investigation. *Southern African Linguistics and Applied Language Studies (SALALS)* 26(2). 255–265.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1).

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune H. B. Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, Grothendieck & Gabor, Green, Peter. 2017. Package 'lme4': Linear mixed-effects models using Eigen and S4. http://CRAN.R-project.org/package = lme4.

Bekker, Ian. 2008. *The vowels of South African English.* Potchefstroom, South Africa: North-West University PhD thesis.

Bekker, Ian. 2013. The Formation of South African English. *English Today* 29(01). 3–9.

Bekker, Ian & Georgina Eley. 2007. An acoustic analysis of White South African English (WSAfE) monophthongs. *Southern African Linguistics & Applied Language Studies* 25(1). 107–114.

Bekker, Ian & Bertus Van Rooy. 2015. The Pronunciation of English in South Africa. In Marnie Reed & John M. Levis (eds.), *The Handbook of English Pronunciation*, 286–300. Chichester: Wiley-Blackwell.

Boersma, Paul & David Weenink. 2014. *Praat: doing phonetics by computer. Version 5.3.63.* http://www.praat.org.

Bowerman, Sean. 2004. White South African English: Phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English: A multimedia reference tool. Vol. 1, Phonology*, 931–942. Berlin: Mouton de Gruyter.

Brato, Thorsten. Manuscript. *The vowels of Botswanan English: An acoustic-phonetic pilot study*.

Brato, Thorsten. 2016. *Variation and Change in Aberdeen English: A Sociophonetic Study.* Frankfurt am Main: Peter Lang.

Brink, Janus D. & Elizabeth C. Botha. 2005. An acoustic comparison of the vowels and diphthongs of first and second language South African English. http://www.prasa.up.ac.za/SpeechProcessingI/PDF/S5JDBrink.pdf.

Bryman, Alan. 2012. *Social research methods,* 4th edn. Oxford: Oxford University Press.

Calenge, Clement. 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecological Modelling*(197). 516-519.

Census 2011. Statistics South Africa. http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf.

Chevalier, Alida. 2019. The reverse short front vowel shift in South African English. In Raymond Hickey (ed.), *English in multilingual South Africa: The linguistics of contact and change*, 151–175. New York, NY: Cambridge University Press.

Clements, Nick & Annie Rialland. 2008. Africa as a phonological area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 36–85. Cambridge: Cambridge University Press.

Coetzee-Van Rooy, Susan. 2012. Flourishing functional multilingualism: evidence from language repertoires in the Vaal Triangle region. *International Journal of the Sociology of Language*(218). 87–119.

Coetzee-Van Rooy, Susan. 2014. The identity issue in bi- and multilingual repertoires in South Africa: Implications for Schneider's Dynamic Model. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The dynamic model and beyond,* 39–57. Amsterdam: Benjamins.

Coetzee-Van Rooy, Susan & Bertus Van Rooy. 2005. South African English: labels, comprehensibility and status. *World Englishes* 24(1). 1–19.

Cook, Susan E. 2009. Street Setwana vs. School Setswana: Language Policy and the Forging of Identities in South Africa. In Jo A. Kleifgen & George C. Bond (eds.), *The languages of*

Africa and the diaspora: *Educating for language awareness*, 96–116. Bristol, UK, Buffalo, NY: Multilingual Matters.

Cruttenden, Alan. 2008. *Gimson's pronunciation of English,* 7th edn. London: Hodder Education.

Da Silva, Arista. 2007. *South African English: A Sociolinguistic Investigation of an Emerging Variety*. Johannesburg: University of the Witwatersrand PhD [Available at http://wiredspace.wits.ac.za/handle/10539/4955].

De Kadt, Elizabeth. 1993. Attitudes towards English in South Africa. *World Englishes* 12(3). 311–324.

De Klerk, Vivian. 1996. Introduction. In Vivian De Klerk (ed.), *Focus on South Africa*, 7–9. Amsterdam: Benjamins.

De Klerk, Vivian. 1999. Black South African English: Where to from here? *World Englishes* 18(3). 311–324.

De Klerk, Vivian & David Gough. 2002. Black South African English. In Rajend Mesthrie (ed.), *Language in South Africa*, 356–378. Cambridge: Cambridge University Press.

Deterding, David. 2006. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association* 36(02). 187.

Deumert, Ana. 2010. Tracking the demographics of (urban) language shift – an analysis of South African census data. *Journal of Multilingual and Multicultural Development* 31(1). 13–35.

Di Paolo, Marianna, Malcah Yaeger-Dror & Alicia B. Wassink. 2011. Analyzing vowels. In Marianna Di Paolo & Malcah Yaeger-Dror (eds.), *Sociophonetics*: *A student's guide*, 87–106. London: Routledge.

DiCanio, Christian. 2013. Visualizing vowel spaces in R: from points to contour maps. http://christiandicanio.blogspot.com/2013/10/visualizing-vowel-spaces-in-r-from.html (5 January, 2019).

Doke, Clement M. 1967. *The Southern Bantu Languages.* London: Dawson of Pall Mall.

Doke, Clement M. & S. M. Mofokeng. 1974. *Textbook of Southern Sotho grammar.* Cape Town: Longman.

Du Plessis, Deon. 2015. *Regionality in White South African English: an acoustic dialectometric investigation.* Potchefstroom, South Africa: North-West University MA thesis.

Du Plessis, Deon. 2019. Regionality in South African English. In Raymond Hickey (ed.), *English in multilingual South Africa: The linguistics of contact and change*, 74–100. New York, NY: Cambridge University Press.

Dubois, Sylvie & Barbara Horvath. 2000. When the music changes, you change too: Gender and language change in Cajun English. *Language Variation and Change* 11(3). 287–313.

ELAN. 2015. *ELAN: (EUDICO Linguistic Annotator)*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands.

Evanini, Keelan & Becky Huang. 2013. Production of English vowels by speakers of Mandarin Chinese with prolonged exposure to English. In, *164th Meeting of the Acoustical Society of America*, 1–9: ASA.

Fabricius, Anne H. 2007. Variation and change in the trap and strut vowels of RP: a real time comparison of five acoustic data sets. *Journal of the International Phonetic Association* 37(03). 293–320.

Field, Andy, Jeremy Miles & Zoë Field. 2012. *Discovering statistics using R.* Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE.

Fought, Carmen. 1999. A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics*(3). 5–23.

Frost, Jim. 2013. The Minitab Blog: Regression Analysis: How to Interpret the Constant (Y Intercept). http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-to-interpret-the-constant-y-intercept.

Frost, Jim. 2014. The Minitab Blog: How High Should R-squared Be in Regression Analysis? http://blog.minitab.com/blog/adventures-in-statistics-2/how-high-should-r-squared-be-in-regression-analysis.

Fuchs, Robert & Ulrike Gut. 2015. An apparent time study of the progressive in Nigerian English. In Peter Collins (ed.), *Grammatical change in English world-wide*, 373–387. Amsterdam: Benjamins.

Gough, David. 1996. Black English in South Africa. In Vivian De Klerk (ed.), *Focus on South Africa*, 53–77. Amsterdam: Benjamins.

Gowlett, Derek. 2006. Zone S. In Derek Nurse & Gérard Philippson (eds.), *The Bantu Languages*, 609–638. New York, NY: Routledge.

Gut, Ulrike. 2004. Nigerian English: phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English: A multimedia reference tool. Vol. 1, Phonology*, 813–830. Berlin: Mouton de Gruyter.

Gut, Ulrike. 2005. Nigerian English prosody. *English World-Wide* 26(2). 153–177.

Hall-Lew, Lauren. 2009. *Ethnicity and Phonetic Variation in a San Francisco Neighbourhood.* Stanford: Stanford University Dissertation.

Harley, Trevor A. 2014. *The psychology of language*: *From data to theory,* 4th edn. London, New York: Psychology Press.

Hartmann, Dieter & Sabine Zerbian. 2009. Rhoticity in Black South African English - a sociolinguistic study. *Southern African Linguistics and Applied Language Studies* 27(2). 135–148.

Hartshorne, Ken. 1992. *Crisis and challenge*: *Black education 1910-1990.* Cape Town, New York: Oxford University Press.

Hay, Jennifer, Katie Drager & Paul Warren. 2009. Careful Who You Talk to: An Effect of Experimenter Identity on the Production of the NEAR/SQUARE Merger in New Zealand English. *Australian Journal of Linguistics* 29(2). 269–285.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4). 458–484.

Hirson, Baruch. 1981. Language control and resistance in South Africa. *African Affairs* 319(80). 219–237.

Hoffmann, Thomas. 2011. The Black Kenyan English vowel system: An acoustic phonetic analysis. *English World-Wide* 32(2). 147–173.

Hothorn, Torsten, Kurt Hornik, Carolin Strobl & Achim Zeileis. 2020. Package 'party'. https://CRAN.R-project.org/package=party.

Huber, Magnus. 2004. Ghanaian English: phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*: *A multimedia reference tool. Vol. 1, Phonology*, 842–865. Berlin: Mouton de Gruyter.

Hundleby, C. E. 1963. *Xhosa-English pronunciation in the South-East Cape.* Grahamstown, South Africa: Rhodes University PhD thesis.

Isiaka, Adeiza L. 2017. *Ebira English in Nigerian Supersystems: Inventory and Variation.* Chemnitz: Technische Universität Chemnitz Dissertation.

Jansen, Sandra. 2017. Change and stability in goose, goat and foot: Back vowel dynamics in Carlisle English. *English Language and Linguistics* 43. 1–29.

Johnson, Daniel E. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3(1). 359–383.

Johnson, Daniel E. 2014. Quantifying Overlap with Bhattacharyya's affinity and other measures! https://danielezrajohnson.shinyapps.io/nwav_44/.

Johnson, Daniel E. 2017. Quantifying Overlap: A Shiny App for NWAV 44 (Chicago, 2014). http://dejonlanguage.blogspot.de/2017/05/quantifying-overlap-shiny-app-for-nwav.html.

Jones, Daniel, Peter Roach, James Hartman & Jane Setter (eds.) (2006). *Cambridge English pronouncing dictionary,* 17th edn. Cambridge: Cambridge University Press.

Kamwangamalu, Nkonko M. 2001. Ethnicity and language crossing in post-apartheid South Africa. *International Journal of the Sociology of Language*(152). 75–95.

Kamwangamalu, Nkonko M. 2004. The language planning situation in South Africa. In Richard B. Baldauf & Robert B. Kaplan (eds.), *Language planning and policy in Africa, Vol. 1*: *Botswana, Malawi, Mozambique and South Africa*, 197–281. Clevedon: Multilingual Matters.

Kamwangamalu, Nkonko M. 2009. South African Englishes. In Braj B. Kachru, Yamuna Kachru & Cecil L. Nelson (eds.), *The handbook of world Englishes*, 158–171. Malden, Mass: Wiley-Blackwell.

Kendall, Tyler. 2009. Vowel Capture Script. http://lingtools.uoregon.edu/scripts/vowel_capture_aug09.praat.

Khabanyane, Khathatso E. 1991. The five vowel heights of Southern Sotho: an acoustic and Phonological Analysis. *Working Papers of the Cornell Phonetics Laboratory*(5). 1–34.

Krug, Manfred & Katrin Sell. 2013. Designing and conducting interviews and questionnaires. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 69–98. Cambridge: Cambridge University Press.

Kruger, Haidee & Bertus Van Rooy. 2017. Editorial practice and the progressive in Black South African English. *World Englishes* 36(1). 20–41.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017a. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13).

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune H. B. Christensen. 2017b. Package 'lmerTests': Tests in Linear Mixed Effects Models. https://cran.r-project.org/web/packages/lmerTest/index.html.

Labov, William. 1984a. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use*: *Readings in sociolinguistics*, 28–53. Englewood Cliffs, N.J: Prentice-Hall.

Labov, William. 1984b. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use*: *Readings in sociolinguistics*, 28–53. Englewood Cliffs, N.J: Prentice-Hall.

Labov, William. 1991. *Sociolinguistic patterns,* 11th edn. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of linguistic change*: *Internal factors.* Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change*: *Social factors.* Oxford: Blackwell.

Labov, William. 2010. *Principles of linguistic change*: *Cognitive and cultural factors.* Chichester, Malden, MA: Wiley-Blackwell.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English*: *Phonetics phonology and sound change; a multimedia reference tool.* Berlin: De Gruyter.

Ladefoged, Peter. 2003. *Phonetic data analysis*: *An introduction to fieldwork and instrumental techniques.* Malden, Mass.: Blackwell.

Lado, Robert. 1957. *Linguistics across cultures*: *Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Lanham, Len W. 1990. Stress and intonation and the intelligibility of South African Black English. In Susan Ramsaran & Alfred C. Gimson (eds.), *Studies in the pronunciation of English*: *A commemorative volume in honour of A.C. Gimson*, 243–260. London: Routledge.

Lanham, Len W. 1996. A History of English in South Africa. In Vivian De Klerk (ed.), *Focus on South Africa,* 19–34. Amsterdam: Benjamins.

Lanham, Len W. & C. A. Macdonald. 1985. *The standard in South African English and its social history,* 2nd edn. Heidelberg: Groos.

Lass, Roger. 1990. A ’standard’ South African vowel system. In Susan Ramsaran & Alfred C. Gimson (eds.), *Studies in the pronunciation of English*: *A commemorative volume in honour of A.C. Gimson*, 272–285. London: Routledge.

Lass, Roger. 2002. South African English. In Rajend Mesthrie (ed.), *Language in South Africa*, 104–126. Cambridge: Cambridge University Press.

Lass, Roger. 2004. South African English. In Raymond Hickey (ed.), *Legacies of colonial English*: *Studies in transported dialects*, 363–386. Cambridge: Cambridge University Press.

Lass, Roger & Susan Wright. 1986. Endogeny Vs. Contact: 'Afrikaans Influence' on South African English. *English World-Wide* 7(2). 201–223.

Lehiste, Ilse & Gordon E. Peterson. 1961. Transitions, Glides, and Diphthongs. *Journal of the Acoustical Society of America* 33(3). 268–277.

Leketi, Makalela. 2004. Making sense of BSAE for linguistic democracy in South Africa. *World Englishes* 23(3). 355–366.

Lenneberg, Eric H. 1967. *Biological foundations of language.* New York: Wiley.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins Publishing Company.

Lewis, M. P. Gary F. Simons & Charles D. Fennig (eds.) (2014). *Ethnologue: Languages of the World: South Africa,* 17th edn. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/country/ZA/status.

Louw, Philippa & Febe de Wet. 2007. The perception and identification of accent in spoken Black South African English. *Southern African Linguistics and Applied Language Studies* 25(1). 91–105.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation.* Prague: Epocha.

Maddieson, Ian. 2006. The Sounds of the Bantu Languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu Languages,* 15–41. New York, NY: Routledge.

Makalela, Leketi. 2004. Making sense of BSAE for linguistic democracy in South Africa. *World Englishes* 23(3). 355–366.

Malan, Karen. 1996. Cape Flats English. In Vivian De Klerk (ed.), *Focus on South Africa*, 125–148. Amsterdam: Benjamins.

Mangiafico, Salvatore S. 2016a. Summary and Analysis of Extension Program Evaluation in R, version 1.13.4: Least Square Means for Multiple Comparisons. http://rcompanion.org/handbook/G_06.html (2 April, 2018).

Mangiafico, Salvatore S. 2016b. Summary and Analysis of Extension Program Evaluation in R, version 1.13.4: What are Least Square Means? https://rcompanion.org/handbook/G_05.html (2 April, 2018).

McArthur, Tom. 2002. *The Oxford guide to world English.* Oxford: Oxford University Press.

Meer, Philipp, Thorsten Brato & José A. Matute Flores. under review. Extending automatic vowel formant extraction to New Englishes: A comparison of different methods.

Mesthrie, Rajend. 2002. Introduction. In Rajend Mesthrie (ed.), *Language in South Africa*, 1–8. Cambridge: Cambridge University Press.

Mesthrie, Rajend. 2004a. Black South African English: morphology and syntax. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English: A multimedia reference tool. Vol. 2, Morphology and syntax*, 962–973. Berlin: Mouton de Gruyter.

Mesthrie, Rajend. 2004b. Synopsis: the phonology of English in Africa and South and Southeast Asia. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive

Upton (eds.), *A handbook of varieties of English: A multimedia reference tool. Vol. 1, Phonology*, 1099–1109. Berlin: Mouton de Gruyter.

Mesthrie, Rajend. 2009a. Contact Linguistics and World Englishes. In Braj B. Kachru, Yamuna Kachru & Cecil L. Nelson (eds.), *The handbook of world Englishes*, 274–288. Malden, Mass: Wiley-Blackwell.

Mesthrie, Rajend. 2009b. Deracialising the GOOSE vowel in South African English: Accelerated linguistic change amongst young, middle class females in post-apartheid South Africa. In Thomas Hoffmann & Lucia Siebers (eds.), *World Englishes - problems, properties and prospects*, 3–18. Amsterdam: Benjamins.

Mesthrie, Rajend. 2010a. Sociolinguistics in South Africa: a critical overview of current research. In Martin J. Ball (ed.), *The Routledge handbook of sociolinguistics around the world*, 187–202. London: Routledge.

Mesthrie, Rajend. 2010b. Socio-phonetics and social change: Deracialisation of the GOOSE vowel in South African English 1. *Journal of Sociolinguistics* 14(1). 3–33.

Mesthrie, Rajend. 2017. Class, gender, and substrate erasure in sociolinguistic change: A sociophonetic study of schwa in deracializing South African English. *Language* 93(2). 314–346.

Mesthrie, Rajend, Alida Chevalier & Timothy Dunne. 2015. A Regional and Social Dialectology of the BATH Vowel in South African English. *Language Variation and Change* 27(1). 1–30.

Mesthrie, Rajend & Jeanne Hromnik. 2011. *Eish, but is it English?: Celebrating the South African variety.* Cape Town: Zebra Press.

Mesthrie, Rajend, Joan Swann, Andrea Deumert & William L. Leap. 2000. *Introducing sociolinguistics.* Edinburgh: Edinburgh University Press.

Milroy, Lesley & Matthew Gordon. 2003. *Sociolinguistics: Method and interpretation.* Malden, Mass.: Blackwell.

Minow, Verena. 2010. *Variation in the grammar of Black South African English.* Frankfurt am Main, Berlin, Bern, Wien: Lang.

Mitchell, A. G. & Arthur Delbridge. 1965. *The pronunciation of English in Australia.* Sydney: Angus and Robertson.

North-West University. 2015. Corporate Profile 2014/2015. http://www.myvirtualpaper.com/doc/north-west-university/nwucorporateprofile_eng_afr_2014-2015/2014100301/#0.

Nycz, Jennifer. 2013. New contrast acquisition: Methodological issues and theoretical implications. *English Language and Linguistics* 17(02). 325–357.

Nycz, Jennifer & Lauren Hall-Lew. 2014. Best practices in measuring vowel merger. *Journal of the Acoustical Society of America* 20. 1–19.

Oehlert, Gary W. 2011. A few words about REML. users.stat.umn.edu/~corbett/classes/5303/REML.pdf.

O'Neill, Mick. 2013. The mathematics of REML. www.stats.net.au/Maths_REML_manual.pdf.

Pebesma, Edzer J. & Roger S. Bivand. 2005. Classes and methods for spatial data in R. *R News* 5(2). 9–13.

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America* 32(6). 693–703.

Pongweni, Alec J. C. 1990. The pronunciation of English by Shona speakers: problems and causes. In Susan Ramsaran & Alfred C. Gimson (eds.), *Studies in the pronunciation of English: A commemorative volume in honour of A.C. Gimson*, 231–242. London: Routledge.

R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. Version 0.98.953. https://www.r-project.org/.

Rasinger, Sebastian M. 2013. *Quantitative research in linguistics: An introduction,* 2nd edn. London: Bloomsbury Academic.

Robins, Richard W. R. C. Fraley & Robert F. Krueger. 2010. *Handbook of research methods in personality psychology.* New York, London: Guilford.

Schiel, Florian. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In, *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1–4.

Schiel, Florian, Andreas Kipp, Thomas Kisler & Ines Wendler. n.d. *MAUS (The Munich Automatic Segmentation System):* Ludwig-Maximilians-Universität München.

Schilling, Natalie. 2013. Investigating Stylistic Variation. In Jack K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, 2nd edn. 327–349. Malden, MA: Wiley-Blackwell.

Schilling-Estes, Natalie. 2007. Sociolinguistic fieldwork. In Robert Bayley & Ceil Lucas (eds.), *Sociolinguistic variation: Theories methods and applications*, 165–189. Cambridge: Cambridge University Press.

Schmied, Josef. 1991. *English in Africa: An introduction.* London: Longman.

Schmied, Josef. 2004. East African English (Kenya, Uganda, Tanzania): Phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A* *handbook of varieties of English: A multimedia reference tool. Vol. 1, Phonology*, 918–930. Berlin: Mouton de Gruyter.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world.* Cambridge, New York, NY: Cambridge University Press.

Schneider, Edgar W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33(1). 9–32.

Schneider, Edgar W. 2019. South Africa in the linguistic modelling of World Englishes. In Raymond Hickey (ed.), *English in multilingual South Africa: The linguistics of contact and change*. Cambridge: Cambridge University Press.

Sharbawi, Salbrina H. 2006. The vowels of Brunei English: An acoustic investigation 27(3). 247–264.

Siebers, Lucia. 2012. *Morphosyntax in Black South African English: A sociolinguistic analysis of Xhosa English.* Tübingen: Narr Francke.

Simo Bobda, Augustin. 2000a. Comparing some phonological features across African accents of English. *English Studies*(3). 249–266.

Simo Bobda, Augustin. 2000b. English pronunciation in sub-Saharan Africa as illustrated by the NURSE vowel. *English Today* 16(4). 41–48.

Smit, Ute. 1996. South African English in the 1990s: A Field study on status, roles and attitudes. *English World-Wide* 17(1). 77–109.

Spencer, Brenda. 2010. Linguistic penetration at Schneider's Phase 4 : acceptability ratings of entrenched features of Black South African English by South Africans outside the originating culture of the variety. *Journal for Language Teaching* 45(2). 133–149.

Strelluf, Christopher. 2016. Overlap among back vowels before /l/ in Kansas City. *Language Variation and Change* 28(3). 379–407.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation.* Cambridge: Cambridge University Press.

Thomas, Erik R. 2011. *Sociophonetics: An introduction.* Basingstoke: Palgrave Macmillan.

Thomas, Erik R. & Tyler Kendall. 2007–2015. NORM: The vowel normalization and plotting suite. http://lingtools.uoregon.edu/norm/index.php (last checked 9 August, 2016).

Torgersen, Eivind & Paul Kerswill. 2004. Internal and external motivation in phonetic change: Dialect levelling outcomes for an English vowel shift. *Journal of Sociolinguistics* 8(1). 23–53.

Trudgill, Peter. 1974. *The social differentiation of English in Norwich.* Cambridge: Cambridge University Press.

Tucker, Archibald N. 1929. *The comparative phonetics of the Suto-Chuana group of Bantu languages.* London: Longman.

Van Rooy, Bertus. 2000. The consonants of BSAE: Current knowledge and future prospects. *South African Journal of Linguistics* 18(sup38). 35–54.

Van Rooy, Bertus. 2002. Stress placement in Tswana English: the makings of a coherent system. *World Englishes* 21(1). 145–160.

Van Rooy, Bertus. 2004. Black South African English: Phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English: A multimedia reference tool. Vol. 1, Phonology*, 943–952. Berlin: Mouton de Gruyter.

Van Rooy, Bertus. 2007. Consonant Clusters and Resyllabification in Black South African English. *Language Matters* 38(1). 26-.

Van Rooy, Bertus. 2014. Convergence and endonormativity at Phase Four of the Dynamic Model. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The dynamic model and beyond*, 21–38. Amsterdam: Benjamins.

Van Rooy, Bertus. 2020. English in Africa. In Daniel Schreier, Edgar W. Schneider & Marianne Hundt (eds.), *The Cambridge handbook of world Englishes*, 210–235. Cambridge: Cambridge University Press.

Van Rooy, Bertus & Lize Terblanche. 2010. Complexity in word-formation processes in New Varieties of South African English. *Southern African Linguistics and Applied Language Studies* 28(4). 357–374.

Van Rooy, Bertus & Gerhard Van Huyssteen. 2000. The vowels of BSAE: Current knowledge and future prospects. *South African Journal of Linguistics* 18(sup38). 15–33.

Van Rooy, Bertus, Sanet Van Rooyen & Herman Van Wyk. 2000. An Assessment of High School Pupils' Attitudes Towards the Pronunciation of Black South African English. *South African Journal of Linguistics* 18(sup38). 187–213.

Wassink, Alicia B. 1999. *A sociophonetic study of Jamaican vowels.* Ann Arbor: University of Michigan Dissertation.

Wassink, Alicia B. 2006. A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. *The Journal of the Acoustical Society of America* 119(4). 2334–2350.

Watermeyer, Susan. 1996. Afrikaans English. In Vivian De Klerk (ed.), *Focus on South Africa*, 99–124. Amsterdam: Benjamins.

Watt, Dominic & Anne H. Fabricius. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1~F2 plane. In D. Nelson (ed.), *Leeds Working Papers in Linguistics and Phonetics 9*, 159–173.

Wells, John C. 1982a. *Accents of English 1*: *An Introduction.* Cambridge: Cambridge University Press.

Wells, John C. 1982b. *Accents of English 3*: *Beyond the British Isles.* Cambridge: Cambridge University Press.

Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Dafydd Gibbon, R. Moore & R. Winski (eds.), *Handbook of standards and resources for spoken language systems*, Part IV, section B. Berlin, New York: Mouton de Gruyter.

Winter, Bodo. 2013. *Linear models and linear mixed effects models in R with linguistic applications.*

Wissing, Daan. 2002. Black South African English: A new English? Observations from a phonetic viewpoint. *World Englishes* 21(1). 129–144.

Wright, Laurence. 1996. The Standardisation Question in Black South African English. In Vivian De Klerk (ed.), *Focus on South Africa*, 149–162. Amsterdam: Benjamins.

Zerbian, Sabine & Etienne Barnard. 2008. Phonetics of intonation in South African Bantu languages. *Southern African Linguistics and Applied Language Studies* 26(2). 235–254.

Appendix

A-1 Overview of speaker metadata

Speaker	Age	Age group	Gender	L1	Home province	Highest education
tyf04	19	young	female	Tswana	Gauteng	CS[13]
syf05	19	young	female	S. Sotho	Gauteng	CS
syf07	19	young	female	S. Sotho	Gauteng	CS
syf06	19	young	female	S. Sotho	Gauteng	CS
sym02	20	young	male	S. Sotho	Free State	CS
zym04	20	young	male	Zulu	Gauteng	CS
xyf03	20	young	female	Xhosa	Gauteng	CS
xym05	20	young	male	Xhosa	North West	CS
oyf13	21	young	female	Tsonga	North West	CS
zym06	21	young	male	Zulu	Gauteng	PS[14]
xym01	21	young	male	Xhosa	Gauteng	CS
xyf08	21	young	female	Xhosa	Gauteng	CS
syf09	21	young	female	S. Sotho	Gauteng	CS
tyf01	21	young	female	Tswana	North West	PS
sym03	21	young	male	S. Sotho	Gauteng	CS
tym07	22	young	male	Tswana	Gauteng	CS
zym09	22	young	male	Zulu	Gauteng	PS
syf10	22	young	female	S. Sotho	Gauteng	CS
xyf11	22	young	female	Xhosa	Gauteng	CS
sym08	23	young	male	S. Sotho	Free State	PS
smm10	25	middle	male	S. Sotho	Gauteng	CS
zmf02	26	middle	female	Zulu	Gauteng	PG[15]
zmf12	27	middle	female	Zulu	Gauteng	PS

[13] CS = completed secondary education, e.g. Bachelor student or non-academic occupation

[14] PS = post-secondary education: completed Bachelor programme or National Diploma

[15] PG = post-graduate education: completed at least an Honours programme

Speaker	Age	Age group	Gender	L1	Home province	Highest education
smm06	30	middle	male	S. Sotho	Gauteng	CS
xmm09	30	middle	male	Xhosa	Gauteng	PS
tmf03	32	middle	female	Tswana	Gauteng	PS
smm01	35	middle	male	S. Sotho	Gauteng	CS
tmm10	35	middle	male	Tswana	Gauteng	PG
smf06	38	middle	female	S. Sotho	Gauteng	PG
smm03	41	middle	male	S. Sotho	Gauteng	PS
smf05	41	middle	female	S. Sotho	Gauteng	PS
smm11	42	middle	male	S. Sotho	Gauteng	PG
nmf02	44	middle	female	N. Sotho	Gauteng	PS
tmm12	47	middle	male	Tswana	Gauteng	PG
smm04	50	middle	male	S. Sotho	Gauteng	PG
smm13	50	middle	male	S. Sotho	Gauteng	PG
xmf04	54	middle	female	Xhosa	Gauteng	PG
dof06	58	older	female	Ndebele	Gauteng	CS
tof01	59	older	female	Tswana	North West	CS
sof05	62	older	female	S. Sotho	Gauteng	CS
vom01	63	older	male	Venda	Gauteng	PG
sof02	66	older	female	S. Sotho	Free State	PS
sof03	71	older	female	S. Sotho	Free State	CS
som02	84	older	male	S. Sotho	Gauteng	PG

A-2 Questionnaire

Place and date of recording:

Participant code:

1 BIO-DATA

1.1 Your age: ………

1.2 Your gender: Female ▢ Male ▢

2 PLACE OF RESIDENCE

2.1 In which province do you currently live? ....

2.2 Were you born in South Africa? Yes: ▢ No: ▢

2.3 Have you stayed abroad for more than six months? Yes: ▢ No: ▢

3 QUALIFICATIONS / PROFESSIONAL CAREER

3.1 If you are a university student, are you a …? (Please tick)

Bachelor: ▢ Honours: ▢ Master's: ▢ PhD: ▢

3.2 What is the highest qualification/level of education you obtained?

…………………………………………………………………………………………

3.3 What is your current occupation/job title (if applicable)?

…………………………………………………………………………………………

4 LANGUAGE SKILLS

4.1 What was the first language you acquired? (Please tick the box)

Zulu	▢	Southern Sotho	▢
Xhosa	▢	Ndebele	▢
Venda	▢	Northern Sotho/Sepedi	▢
Tswana	▢	English	▢
Tsonga	▢	Afrikaans	▢
Swati	▢	Other, please specify: ……………………………..	

4.2 What other South African languages did you acquire? (Please tick the box)

Zulu	▢	Southern Sotho	▢
Xhosa	▢	Ndebele	▢
Venda	▢	Northern Sotho/Sepedi	▢
Tswana	▢	English	▢
Tsonga	▢	Afrikaans	▢
Swati	▢	Other, please specify: ……………………………..	

4.3 Which language do you speak best? (Please tick the box)

Zulu	▢	Southern Sotho	▢
Xhosa	▢	Ndebele	▢
Venda	▢	Northern Sotho/Sepedi	▢
Tswana	▢	English	▢
Tsonga	▢	Afrikaans	▢
Swati	▢	Other, please specify: ……………………………..	

4.4 How often do you speak English in the following contexts? (Please tick)

	in your family	with your friends	at the market / in town	at work	at university
always					
frequently					
sometimes					
rarely					
never					

5 LANGUAGE HABITS

5.1 In your opinion, which of the following languages is (or are) the key for power and success? Please check any that apply.

Zulu ▢ Southern Sotho ▢

Xhosa ▢ Ndebele ▢

Venda ▢ Northern Sotho/Sepedi ▢

Tswana ▢ English ▢

Tsonga ▢ Afrikaans ▢

Swati ▢ Other, please specify: ……………………………..

5.2 How well do you agree with the following statements? (Please tick)

	I like to speak English	I feel confident using English	It is important to be good at English	English forms part of my identity
Agree				
Rather agree				
Rather disagree				
Disagree				

THANK YOU VERY MUCH!

A-3 Reading Passage

The boy who cried 'wolf'

There was once a poor shepherd boy who used to watch his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting 'Wolf, Wolf.' As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety, and two of his cousins even stayed with him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However, not long after, a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before. Unfortunately, as all the villagers were convinced that he was trying to fool them a third time, they told him 'Go away and don't bother us again.' And so the wolf had a feast.

A-4 Wordlist

author – ghost – use – tourist – chubby – palm – got – flask – architecture – explosion – sister – log – foot – badge – goat – thunder – structure – disaster – measure – juice – hug – zoo – shy – boy – goose – during – shine – choose – thousand – sheer – thought – sign – lodge – theft – gut – scratchy – bridges – heart – achieve – face – league – short – hedge – start – force – about – humble – nation – mouth – huge – leak – noise – sure – fragile – ship – poor – kit – dress – couch – seizure – pig – cheers – lock – judge – cure – hatch – cloth – fuse – stretch – shark – corner – guard – fleece – letter – zebra – happy – adventure – gaze – coast – seat – oblige – horses – cut – genre – page – bath – jungle – comma – north – pure – much – card – teach – pick – commercial – choice – sigh – shoot – church – nurse – chase – lot – beer – allergic – bear – sit – peach – kisses – cheat – neighbourhood – strut – dummy – square – price – howl – nature – trap – shop – chime – adjective – possible – chess – near – pitch – cot