
**GLOBALISATION
VERSUS
INTERNAL DEVELOPMENT:
THE REVERSE SHORT FRONT VOWEL SHIFT
IN SOUTH AFRICAN ENGLISH**

ALIDA CHEVALIER

Supervisor: Professor Rajend Mesthrie



Thesis presented for the Degree of

Doctor of Philosophy

in Linguistics

School of African & Gender Studies, Anthropology & Linguistics

Faculty of Humanities

University of Cape Town

May 2016

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

ABSTRACT

Globalisation versus Internal Development: the Reverse Short Front Vowel Shift in South African English.

Alida Chevalier | May 2016

The South African Chain Shift involved the raising of the short front vowels KIT, DRESS and TRAP when compared to Received Pronunciation (Lass & Wright 1986). This raising was particularly evident in the speech of middle class white speakers of South African English, as well as coloured speakers in the Cape. Recent scholarship has suggested that this raising is being reversed in the speech of young white South Africans. In particular, Bekker and Eley (2007) and Bekker (2009) report the lowering and retraction of TRAP. Mesthrie (2012a) reports not only the lowering and retraction of TRAP, but also the lowering of KIT and DRESS. In addition, scholars such as Mesthrie (2010) have found post-segregation deracialisation of middle class South African English.

This thesis therefore investigates the extent to which the reversal of the older South African Chain Shift exists in the speech of white and black middle class South Africans from Cape Town. It furthermore explores the potential merger between TRAP~STRUT and KIT~DRESS. In so doing, 53 participants in sociolinguistic interviews are reported on. The Forced Alignment and Vowel Extraction Toolkit was utilised for formant measurement and extraction. Statistical testing via R was performed, including linear mixed-effects modelling, random forest analyses, conditional inference trees, Euclidean Distance measures, Welch's Two Sample *t*-tests and Pillai Scores.

The analysis finds evidence of the reversal of the South African Chain Shift in the speech of participants under the age of 30. In particular, speakers aged between 18 and 25 participate the most in lowering KIT, DRESS and TRAP. Moreover, the short front vowels are retracting in the speech of younger Capetonians, indicating that within the process of vowel lowering, further innovation occurs via vowel retraction. The Reverse Vowel Shift is found to be a combination of push and pull chains: the fronting of FOOT causes the lowering of KIT, and the lowering of TRAP causes the lowering of DRESS. The retraction of TRAP furthermore causes the backing and raising of STRUT, such that an anti-clockwise rotation of the short front vowels (barring LOT) is evidenced in South African English.

The Reverse Vowel Shift evident in Cape Town is similar to trends observed in California, Canada, southeast England, Ireland and Australia. This illustrates the effects of globalisation on English in South Africa, though internal motivations are also responsible.

ACKNOWLEDGEMENTS

I am so grateful for the blessings in my life, and first and foremost I thank the Lord God for carrying me through this process and for surrounding me with an incredible support system. I feel so blessed to be able to produce a doctoral dissertation and to have so many people involved in the process.

I'd like to acknowledge Prof Rajend Mesthrie's research chair in Migration, Language and Social Change, via the National Research Foundation's South African Research Chairs Initiative (grant no. 64805). This funding has supported my work enormously, for which I am very grateful. The support I received from the University of Cape Town Staff PhD Bursary Scheme in 2015 is also much appreciated.

The University of Cape Town's Humanities Faculty Research Ethics Committee granted me access to the student body. In particular I'd like to thank Mastin Prinsloo, Robyn Udemans and Fiona Ross for their very quick response to my application.

I received tremendous technical support from Kabelo Rameste, Ingrid Rosenfelder, Josef Fruehwald, Thomas Hoffman, Harald Baayen, Martin Hilpert, Christopher Strelluf, Greg Duckworth, Mark Fairbrother, Carl Blom, Ayesha Abrahams, Stephen Browne, Yolandi Ribbens-Klein, Nadine Volmink, and Lauren Hall-Lew. Without them I would have taken three years just to program my MAC, install FAVE, learn R, understand Euclidean Distances/Pillai Scores and transcribe. I am eternally grateful for their eager willingness to help and to answer my frequent questions. A special thank you to Laurel MacKenzie for sending me the British dictionary and for her willingness to talk with me about it. A special thank you also to Joe Fruehwald and the other participants in the FAVE users group for always helping me with codes and solutions.

The analysis and interpretation of my data was helped along enormously by the input I received at NWA43 in October 2014. Many thanks to those who took the time to talk with me about my research, particularly Raymond Hickey and James Grama for their valuable suggestions. I would also like to thank Sali Tagliamonte, Matt Hunt-Gardiner and Derek

Denis for the excellent workshop they ran on statistical methods (“Quantitative methods: new trends and perspectives”).

No sociolinguistic study is possible without speakers. Thank you to all of you for spending an hour or more with me. I can’t wait to read all the novels you have recommended. Thank you to Aléz Odendaal, Liza King, Chiko Chamanga and Bruce Wileman for putting people in touch with me. Thanks also to Raj for letting me borrow a few of his.

On a more personal note:

Rajend Mesthrie, supervisor and mentor: over the past 7 years you have taught, supported and guided me through three dissertations. Thank you for every single correction, suggestion and criticism. Thank you for your unwavering support, both financially and intellectually. Thank you for sending me to conferences. Thank you for the unlimited access to your personal library, which seems to have largely migrated to my office. Thanks to your bibliophilia I had access to every book I needed! Thank you also for your patience in dealing with my relative inexperience in extended thesis writing. I could not have asked for a better person to guide me through this process.

Tracey Toefy: friend, peer and iMessage correspondent. Without your tenacity I would still be using PRAAT manually. Thank you for reading my thesis and for commenting on it. I can’t tell you what it means to me to have you as a sounding board. Thank you for pushing me to learn R, and for sharing your codes with me. Thank you for supporting me through all my years of postgraduate study.

Lobke Minter: fellow #academicintraining, impostor, bookworm and friend. You have supported me more than you know. Thank you for reading and editing my thesis. (Any remaining mistakes are my own.) Thank you for never making me feel like the only crazy person in the world. Thank you for always going out of your way for me. Thank you for everything!

Karen, Carl, Mark, Megan, Lindy. Thank you for being there to celebrate every milestone with me, and for picking me up after every failure. I could not have asked for better friends. A special thanks to Karen, Megan and Lindy for all the supportive whatsapps during the final stages of writing.

My family: Jacobs Seniors, Jacobs Juniors, Turners, Chevalier Seniors, Buissinnes and Korstens. You are all so very precious to me. I would never have made it through anything were it not for your unwavering support and prayers. Thank you for all that you do for me and for all that you mean to me. I love each and every one of you more than you know. A special thank you to my two mothers, my mom Martie and my mom-in-law Noon: your support, prayers and encouragement really brought me through the stages of doubt and struggle, and I am so thankful for that!

Byron: You have been my rock, my comfort and my support for well over a decade. Thank you for always believing in me, especially when I didn't believe in myself. Thank you for rallying the troops when I needed a pick-up. Thank you for being the perfect husband to a dissertating wife. I know it wasn't easy. Thank you for pushing me towards my dream though it meant postponing some of ours. I love you.

DEDICATION

Hierdie verhandeling is vir my ouers, Paul en Martie, and for my husband, Byron.
Ook vir Oupa, omdat hy altyd wou weet hoe dit met my werk gaan, en omdat ek sy
nuuskierigheid geërf het.

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is one's own.
2. I have used the Harvard convention for citation and referencing. Each significant contribution to and quotation in this thesis from the work, or works, of other people has been acknowledged through citation and reference.
3. This thesis is my own work.
4. I have not, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Full Name: Alida Chevalier

Student Number: JCBALI004

Signature:

Signed by candidate

Signature Removed

Date: 20 May 2016

TABLE OF CONTENTS

Abstract.....	i
Acknowledgements	ii
Dedication.....	v
Plagiarism Declaration	vi
Table of Contents.....	vii
List of Figures.....	ix
List of Tables	xiv
List of Abbreviations	xv
Chapter 1: Introduction.....	1
1.1 Aim and scope of the research	3
1.2 Terminology and labelling.....	4
1.3 The chain shift and merger of vowels.....	5
1.3.1 Definitions and approaches	6
1.3.2 Labov’s principles and patterns for vowel shift and merger.....	14
1.4 Global vowel change patterns.....	19
1.4.1 The United Kingdom.....	20
1.4.2 The Southern Hemisphere.....	22
1.4.3 The Northern Hemisphere.....	27
1.5 Summary and thesis outline	33
Chapter 2: South African English.....	34
2.1 The development of (post)colonial Englishes.....	34
2.2 The socio-historical development of South African English.....	44
2.3 Descriptions of South African English	54
2.3.1 Classic descriptions of the monophthongal system.....	57
2.3.2 The Older Short Front Vowel Shift.....	61
2.3.3 The acoustic era.....	63
2.3.4 Vowel mergers	68
2.3.5 The Reverse Short Front Vowel Shift.....	69
2.4 Conclusion	70
Chapter 3: Research & Analytical Methodology	71
3.1 Data collection	71
3.2 The sample	73
3.3 The interview	75
3.4 Automatic formant measurement.....	78
3.4.1 Installing FAVE	78
3.4.2 FAVE-Align.....	82

3.4.3	FAVE-Extract.....	87
3.4.4	Working with the output	90
3.5	Statistical approaches and graphical presentation.....	91
3.5.1	Statistical testing in R.....	92
3.5.2	Graphical presentations of the statistical and vowel data	104
Chapter 4:	Results: Short Front Vowels	105
4.1	Overview of the vowel space	105
4.2	The KIT /ɪ/ vowel.....	112
4.3	The DRESS /e/ Vowel	119
4.4	The TRAP /æ/ vowel	128
4.5	The Reverse Short Front Vowel Shift.....	137
4.6	Other Capetonians.....	138
4.6.1	KIT	140
4.6.2	DRESS	143
4.6.3	TRAP.....	149
4.7	Conclusion	152
Chapter 5:	Results: other Vowels and Mergers	154
5.1	The STRUT /ʌ/ vowel.....	154
5.2	The FOOT /ʊ/ vowel.....	160
5.3	The LOT /ɒ/ and THOUGHT /ɔ:/ vowels.....	166
5.4	Summation: all short vowels and THOUGHT.....	172
5.5	KIT~DRESS merger	173
5.6	TRAP~STRUT merger	184
5.6.1	Minimal pairs tests	189
5.7	Changes to South African English.....	192
Chapter 6:	Discussion and Concluding Remarks	193
6.1	Short Front Vowel Lowering discussed.....	193
6.1.1	Endogeneity or exogeneity?	197
6.2	Transmission of the Reverse Vowel Shift	202
6.3	Indexicality and Enregisterment	211
6.4	Summary of objectives and findings.....	219
6.5	Significance and impact of the study	222
6.6	Future endeavours	222
6.7	Conclusion	224
Appendices	225
Works Cited	230

LIST OF FIGURES

Figure 1.1: Percentage of English ‘main language’ speakers per population group per census year (Statistics South Africa).	2
Figure 1.2 Chain shifting illustrated, based on Gordon (2011, 784).	7
Figure 1.3: Overlapping trajectories in the Northern Cities Shift (Gordon 2002, 255).	8
Figure 1.4: The [±peripheral] tracks followed by changing vowels.....	14
Figure 1.5: Typical directions of vowel shifts along the [±peripheral] track.	15
Figure 1.6: The Northern Cities Shift as illustrated by Thomas (2011, 282).	17
Figure 1.7: The Southern Shift as illustrated by Thomas (2011, 282).	17
Figure 1.8: The Canadian Shift (Thomas 2011, 283).	18
Figure 1.9: Chain Shift in Ashford, United Kingdom (Torgersen & Kerswill 2004, 40). ..	20
Figure 1.10: Vowel changes in Reading, United Kingdom (Torgersen & Kerswill 2004, 45).	21
Figure 1.11: The New Zealand English Short Front Vowel Shift.....	23
Figure 1.12: The 1990s Dublin Vowel Chain Shift (Hickey 2005, 49).....	27
Figure 1.13: The current chain shift in Dublin English (Hickey 2013).....	28
Figure 1.14: The California Shift among 13 speakers aged 19 to 29	31
Figure 2.1: Map of South Africa’s political components in the 1800s.....	45
Figure 2.2: Bekker’s Three-Stage Koinéisation Model of the development of SAfE (Bekker 2012a, 136).....	49
Figure 3.1: Map of South Africa, showing the various provinces and prominent cities/towns.....	72
Figure 3.2: Map of the suburbs and towns of Cape Town, South Africa.....	72
Figure 3.3: Example of words found in the British English Example Pronouncing dictionary (Robinson 1994).....	81
Figure 3.4: Sample of a transcription exported from ELAN as a tab-delimited file.	82
Figure 3.5: Example of a file of unknown elements in a transcription, created by FAVE-Align.....	83
Figure 3.6: Example of an input file to add unknown elements in a transcription to the dictionary.....	84
Figure 3.7: Selection of a TextGrid created by FAVE Align.	85
Figure 3.8: Example of a misaligned TextGrid for the word <i>booed</i> /bu:d/.	86
Figure 3.9: Example of a corrected TextGrid alignment for the word <i>booed</i> /bu:d/.	86

Figure 3.10: Example of PRAAT’s formant tracking using LPC analysis.	88
Figure 3.11: Example boxplot for the F2 of KIT2 (node 11, table 3.5).....	100
Figure 4.1: Means per speaker of each monophthong (in all environments) in the sample of Capetonian English speakers.	106
Figure 4.2: Means per speaker of each monophthong (in all environments) in the sample of Capetonian English speakers, separated by age.....	107
Figure 4.3: Mean values of all short vowels per speaker with standard deviation (x2) ellipses.....	109
Figure 4.4: Conditional inference tree for F1 (height) of all short vowels.....	110
Figure 4.5. Conditional inference tree for F2 (frontness) of all short vowels.	111
Figure 4.6: Mean values and standard deviation (x2) ellipses for the KIT allophones for all speakers.	113
Figure 4.7: Conditional inference tree for KIT: F2 for ‘environment’, ‘gender’ and ‘date of birth’.....	114
Figure 4.8: Conditional inference tree for KIT: F1 for ‘environment’, ‘gender’ and ‘date of birth’.....	116
Figure 4.9: Mean values and standard deviation (x2) ellipses for DRESS for all speakers by environment.....	120
Figure 4.10: Conditional inference tree for DRESS before /l/: F2 for ‘gender’ and ‘date of birth’.....	122
Figure 4.11: Conditional inference tree for DRESS (men): F1 for ‘environment’ and ‘date of birth’.....	123
Figure 4.12: Conditional inference tree for DRESS (women): F1 for ‘environment’ and ‘date of birth’.....	124
Figure 4.13: Conditional inference tree for DRESS: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1989.	125
Figure 4.14: Conditional inference tree for DRESS: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1989.....	127
Figure 4.15: Mean values and standard deviation (x2) ellipses of the TRAP vowel and its environments.	129
Figure 4.16: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1987.	130
Figure 4.17: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1987.	131

Figure 4.18: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1987.	132
Figure 4.19: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1987.	133
Figure 4.20: Mean values, per speaker, of the TRAP vowel of 20 younger women compared to two older women.	135
Figure 4.21: Schematic representation of the Older Vowel Shift.....	137
Figure 4.22: Schematic representation of the Reverse Vowel Shift.....	138
Figure 4.23: Mean values of KIT, DRESS and TRAP per speaker for young white and black middle class Capetonians.	139
Figure 4.24: Conditional inference tree for KIT: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	140
Figure 4.25: Conditional inference tree for ‘other’ KIT: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	142
Figure 4.26: Conditional inference tree for DRESS_1: F1 for ‘gender’, ‘ethnicity’ and ‘date of birth’.....	143
Figure 4.27: Conditional inference tree for DRESS_1: F2 for ‘gender’, ‘ethnicity’ and ‘date of birth’.....	144
Figure 4.28: Conditional inference tree for ‘other’ DRESS for women: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	145
Figure 4.29: Conditional inference tree for ‘other’ DRESS for men: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	146
Figure 4.30: Conditional inference tree for ‘other’ DRESS for speakers born in or before 1989: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	147
Figure 4.31: Conditional inference tree for ‘other’ DRESS for speakers born after 1989: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.....	148
Figure 4.32: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’ (>1987).	149
Figure 4.33: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’ (>1987).	151
Figure 5.1: Standard deviation (x2) ellipses of the subclasses of the STRUT vowel per environment.....	155
Figure 5.2: Conditional inference tree for STRUT: F1 for ‘environment’, ‘gender’ and ‘date of birth’.....	156

Figure 5.3: Conditional inference tree for STRUT: F2 for ‘environment’, ‘gender’ and ‘date of birth’	158
Figure 5.4: Standard deviation (x2) ellipses of the subclasses of the FOOT vowel per environment.....	161
Figure 5.5: Conditional inference tree analysis for FOOT: F1 for ‘gender’, ‘environment’ and ‘date of birth’	162
Figure 5.6: Conditional inference tree analysis for FOOT: F2 for ‘gender’, ‘environment’ and ‘date of birth’	163
Figure 5.7: Casual style tokens and standard deviation (x2) ellipses for KIT6 and FOOT0.	164
Figure 5.8: Standard deviation (x2) ellipses and mean values of all KIT tokens together with FOOT_0, showing encroachment on KIT’s space.....	165
Figure 5.9: Condition inference tree for LOT for F1: ‘gender’ and ‘date of birth’.	166
Figure 5.10: Condition inference tree for LOT for F2: ‘gender’ and ‘date of birth’.	167
Figure 5.10a: Mean values of all short vowels per speaker with standard deviation (x2) ellipses.....	169
Figure 5.11: Conditional inference tree for THOUGHT for F1: ‘gender’ and ‘date of birth’.	171
Figure 5.12: Standard deviation (x2) ellipses illustrating the overlap between KIT and DRESS in all environments.	173
Figure 5.13: Means and standard deviation (x1) of the short vowels of 27 young women in word list style (source: Bekker 2009,197).	174
Figure 5.14: Conditional inference tree for retracted KIT and DRESS_m: F2 for ‘gender’, ‘environment’, ‘vowel’ and ‘date of birth’	175
Figure 5.15: Conditional inference tree KIT2 & DRESS (excluding /l/) F2: ‘gender’, ‘environment’ and ‘date of birth’	176
Figure 5.16: Very close KIT and DRESS tokens of 4 speakers with Pillai Scores under 0.1.	178
Figure 5.17: Distinct KIT and DRESS of 2 speakers with Pillai Scores over 0.5.....	180
Figure 5.18: Close KIT and DRESS for 38 speakers with Pillai Scores between 0.1 and 0.5.	181
Figure 5.19: Conditioned distance between KIT and DRESS: all tokens (left) and standard deviation (x2) ellipses showing overlap (right).	183

Figure 5.20: All casual style tokens of TRAP and STRUT with Pillai Scores between 0.2 and 0.5.....	186
Figure 5.21: Standard deviation (x2) ellipses of all casual style tokens of TRAP and STRUT with Pillai Scores between 0.2 and 0.5.	186
Figure 5.22: Casual style tokens of TRAP and STRUT with Pillai Scores between 0.5 and 0.8.....	187
Figure 5.23: Standard deviation (x2) of all casual style tokens of TRAP and STRUT with Pillai Scores between 0.5 and 0.8.....	188
Figure 5.24: Conditional inference trees: TRAP and STRUT for F1 and F2 with ‘vowel’ as the predictor.	189
Figure 5.25: Minimal pairs of TRAP and STRUT tokens for 21 selected speakers.	190
Figure 5.26: Vowel changes in operation in South African English.	192
Figure 6.1: Comparison of 6 adult women (a) to 4 pre-adolescent girls (p) for KIT, DRESS and TRAP.	203
Figure 6.2: Conditional inference tree for KIT F2 by ‘environment’ and ‘date of birth’: children versus adults.	204
Figure 6.3: Conditional inference tree for six women, showing the KIT split along F2. .	205
Figure 6.4: Conditional inference tree for four children, showing the KIT split along F2.	206
Figure 6.5: Conditional inference tree for DRESS F1 by ‘environment’ and ‘date of birth’: children versus adults.....	207
Figure 6.6: Conditional inference tree for DRESS F2 by ‘environment’ and ‘date of birth’: children versus adults.....	207
Figure 6.7: Conditional inference tree for TRAP F1 by ‘environment’ and ‘date of birth’: children versus adults.....	208
Figure 6.8: Conditional inference tree for TRAP F2 by ‘environment’ and ‘date of birth’: children versus adults.....	209
Figure 6.9: Responses on YouTube to an advertisement by <i>Capitec</i> Bank.	216

LIST OF TABLES

Table 1.1: Number of speakers of English as a ‘main language’ in South Africa per census year (Statistics South Africa).	1
Table 1.2: Grimm’s Law, as presented by Hock (1991, 37).	6
Table 1.3: Principles governing chain shift and merger.	16
Table 1.4: Labov’s chain shift patterns.	16
Table 1.5: Diachronic and synchronic vowel changes in Australian English	25
Table 1.6: Vowel change in Australian English over 100 years	26
Table 2.1: Comparison between KIT, DRESS and TRAP in SAfE and RP.	61
Table 3.1: Demographic information of the selected sample of 53 speakers.	75
Table 3.2: The ARPAbet symbols of transcription with corresponding lexical sets (source: http://fave.ling.upenn.edu/usingFAAValign.html).	80
Table 3.3: Example of R codes and results for a linear mixed-effects regression model.	94
Table 3.4: Example of R codes and results for a linear mixed-effects regression model, including factor interactions.	95
Table 3.5: Example of R codes and results for a random forest model.	97
Table 3.6: Example of R codes and results for a conditional inference tree.	99
Table 3.7: Example of Welch’s Two Sample <i>t</i> -test in R for the F1 [4] and F2 [5] of ChloeV’s TRAP and STRUT vowels.	101
Table 3.8: Example of a Pillai Score generation in R for TRAP and STRUT [2] per speaker for F1 [5] and F2 [6].	103
Table 4.1: The allophones of KIT present in the data.	112
Table 4.2: Summary of changes to KIT in SAfE.	117
Table 4.3: Number of DRESS tokens per environment.	119
Table 4.4: Mixed-effects regression results for DRESS.	121
Table 4.5: Number of tokens of TRAP by environment.	128
Table 5.1: Number of tokens of STRUT by environment.	154
Table 5.2: Number of tokens of FOOT by environment.	160
Table 5.3: Distance measures between close KIT and DRESS.	178
Table 5.4: Comparison of KIT and DRESS: results of Welch’s <i>t</i> -tests.	179
Table 5.5: Distance measures between distinct KIT and DRESS.	180
Table 5.6: Distance measure for KIT and DRESS Set 3 for 3 speakers.	181
Table 5.7: Distance and significance measures for the closest tokens of casual style TRAP and STRUT.	185
Table 5.8: Pillai Scores and Euclidean Distances for TRAP/STRUT minimal pairs.	191

LIST OF ABBREVIATIONS

Set A: General abbreviations.

ANDOSL	Australian National Database of Spoken Language
ANOVA	Analysis of Variance
BSAfE	Black South African English
cforest	Random Forest
ctree(s)	Conditional inference tree(s)
ETE(s)	Extraterritorial English(es)
F1	Formant 1
F2	Formant 2
FAVE	Forced Alignment and Vowel Extraction
Hz	Hertz
L1	First Language
L2	Second Language
lmer	Mixed-Effects Regression Model using {lmerTest}
MANOVA	Multivariate Analysis of Variation
ONZE	Origins of New Zealand English Project
RP	Received Pronunciation
SAfE	South African English
WSAfE	White South African English

Set B: Abbreviations for fixed-effect predictors and phonetic environments.

O	A vowel in the free position
ak	STRUT occurring after /k/
al	STRUT occurring after /l/
am	STRUT occurring after /m/
an	STRUT occurring after /n/
B	black
bg	STRUT occurring before /b/
bk	STRUT occurring before /k/
bl	STRUT occurring before /l/
bm	STRUT occurring before /m/
bn	STRUT occurring before /n/
G	female (to avoid confusion with F for ‘formant’)
g	Vowel occurring before /g/
k	Vowel occurring before /k/
KIT0	The KIT vowel occurring in the free position
KIT2	The KIT vowel occurring in word initially, after /h/ and before/after velars
KIT3	The KIT vowel occurring before or after palato-alveolars
KIT5	The KIT vowel occurring after /l/ and /r/, and before or after bilabials
KIT6	The KIT vowel occurring after /w/ or before /l/
l	Vowel occurring before /l/
M	male
m	Vowel occurring before /m/
n	Vowel occurring before /n/
ng	Vowel occurring before /ŋ/
r	STRUT occurring after /r/
W	white

CHAPTER 1: INTRODUCTION

The variationist approach pioneered by Labov rejects the traditional belief that linguistic changes can only be observed after the fact. Indeed, decades of research into this paradigm have shown the value of studying language change in progress, and the benefits of this approach are perhaps nowhere more evident than in the re-examination of traditional concepts like chain shifts and mergers
Gordon 2002, 245

Nearly 5 million South Africans speak English, which, after isiZulu, isiXhosa and Afrikaans, ranks fourth on the list of languages used as a ‘main language’ in South Africa (Census 2011, Statistics South Africa).¹ As a result of apartheid policy (1948-1994), varieties of English in South Africa developed under segregated conditions, leading to ethnically identifiable varieties of South African English (SAfE). In democratic South Africa, English is often seen as a language of access to education, technology and the global market. Many parents raise and educate their children in English with the vision of bettering their chances of success, often at the expense of the familial mother tongue (see De Klerk 2000).

This attitude towards the status and importance of English perhaps partially explains the increase in the number of L1 English speakers in South Africa over the past 15 or so years. Table 1.1 shows a comparison of census data collected in 1996, 2001 and 2011 regarding the main language of South Africans. Of particular interest is the substantial increase seen for the black population, who are traditionally not speakers of English as a main language. In only 15 years the number has increased by well over a million, which is substantially more than the other population groups (more traditionally associated with speaking English), who increase by around 400,000 at most.

Population Group	1996 Census	2001 Census	2011 Census
Black African	113,132	183,631	1,167,913
Coloured	584,101	756,076	945,847
Indian or Asian	974,654	1,045,845	1,094,317
White	1,711,603	1,687,661	1,603,575
Total	3,383,490	3,673,213	4,811,652

Table 1.1: Number of speakers of English as a ‘main language’ in South Africa per census year (Statistics South Africa).

¹ ‘Main language’ is the term used by the census. I retain the use of it since it is unclear whether it means ‘first language’ or ‘most used language’, though it is likely ‘most used’.

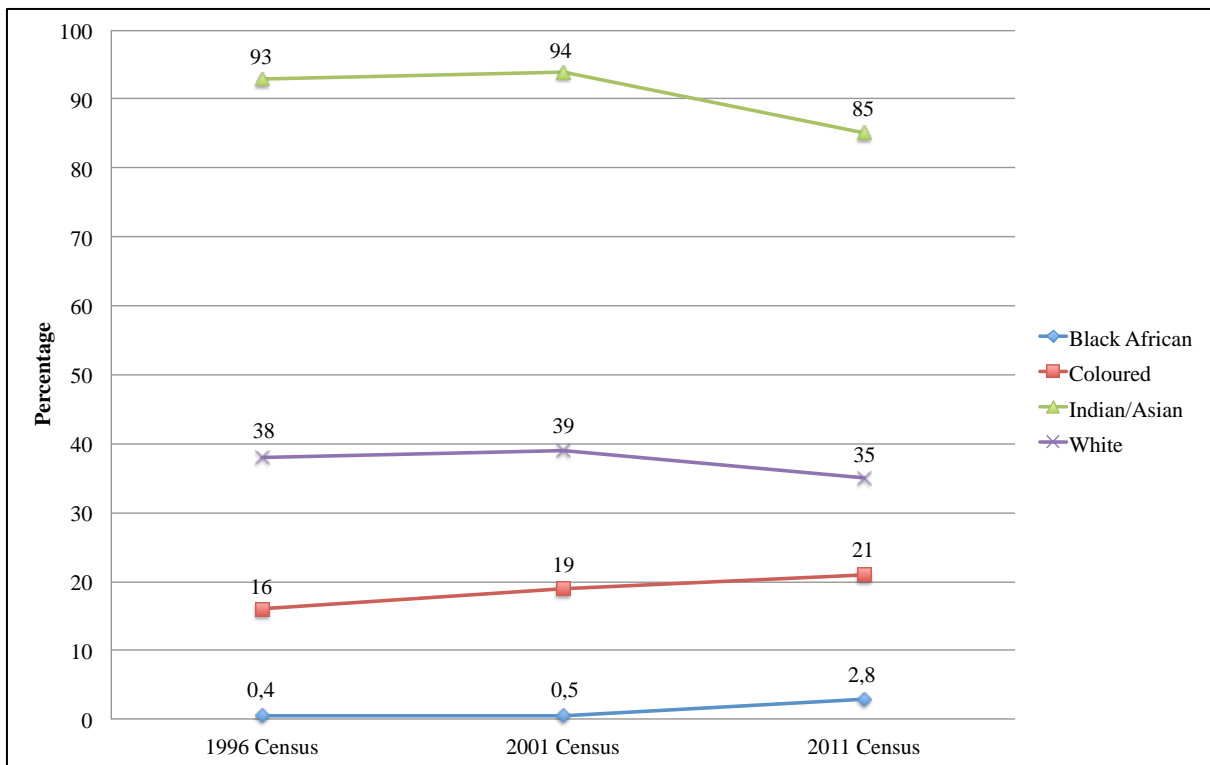


Figure 1.1: Percentage of English ‘main language’ speakers per population group per census year (Statistics South Africa).

Figure 1.1 shows the speakers of English as a main language as a percentage of the overall size of the population group for each ethnicity. It confirms, in particular, a large increase in the proportion of English-using black South Africans, which more than doubles between 2001 and 2011, though the proportion is very small at 2.8%. The other groups show a marginal increase (by 2% for coloured South Africans) and a decrease in the proportion of English users by 9% for Indian/Asian South Africans, and 4% for white South Africans. Traditionally, coloured and white South Africans are also associated with using Afrikaans, while the decrease in the use of English in the Indian group possibly indicates increased use of Indian languages, or the recent immigration of Indians.

Mesthrie (2010, 5) posits the social segregation policies of apartheid and its collapse as creating “near-laboratory conditions for the study of speech divergence and subsequent convergence.” Varieties of English spoken by young, middle class people are becoming different from the varieties of their parents. Various studies document these changes (such as Da Silva 2007, Mesthrie 2010, Morreira 2012, Wilmot 2014 and Toefy 2014), showing that social class, identity and social networks play a role in the variety spoken over and above traditional ethnicity.

This thesis investigates the possibility of two changes currently in progress in the English of young South Africans: the lowering of the short front vowels /ɪ/, /e/ and /æ/ (Bekker 2009; Mesthrie 2012a) and the potential merger of /æ/ and /ʌ/. To firmly articulate the aims of this thesis with regards to these changes, the research questions are provided in section 1.1. A note on terminology is provided in section 1.2. Section 1.3 provides an overview of chain shifts and mergers and theoretical approaches to their study. Section 1.4 outlines global vowel chain shifts and mergers as a starting point of comparison to SAfE that follows in Chapter 2. Section 1.5 briefly summarises this chapter and provides a breakdown of the chapters to follow.

1.1 AIM AND SCOPE OF THE RESEARCH

Lass and Wright (1986) document the South African Chain Shift (henceforth Older Short Front Vowel Shift), which involves the raising of the /ɪ/, /e/ and /æ/ vowels in the English of white South Africans.² Bekker's (2009) study, based on the English of white South African females, provides preliminary evidence that the Older Short Front Vowel Shift is undergoing change: the vowels that were raised are now lowering, comprising a reversal of the older shift. I term this the Reverse Short Front Vowel Shift. Bekker (2009) also found evidence that /ʌ/ is fronting while /æ/ is backing and lowering. He predicts that this could culminate in one of two things: either /æ/ ~ /ʌ/ merger, or a repositioning resulting in /æ/ being lower than /ʌ/ where it was formerly higher (Bekker 2009, 204).

Auditory observations indicate that STRUT is very similar to TRAP in certain varieties of South African English (SAfE). For example, the word *plug* /plʌg/ sounds like *plæg* /plæg/, and British singer James Blunt's surname is variably realised as /blænt/.³ Though the Reverse Short Front Vowel Shift predicts /æ/ lowering, I also investigate the possibility of /æ/ ~ /ʌ/ merger given these auditory perceptions in addition to the findings and subsequent predictions of Bekker (2009) and Bekker and Eley (2007).

In documenting post-apartheid sociolinguistic change, Mesthrie's (2010) study of /u:/ fronting in SAfE showed that, in the middle classes, black and white speakers were the most similar in their degree of /u:/ fronting, followed by Indian speakers and then coloured speakers. Based on these findings, he suggests that SAfE is being 'deracialised': middle class

² This shift is discussed in detail in Chapter 2, section 2.3.5.

³ Tim Thabethe, a radio host on Cape Town's KFM was heard saying this on 11 July 2014.

English varieties in South Africa are now becoming based on something other than ethnicity (Mesthrie 2010, 28). In light of this, the participation or lack of participation in the Reverse Short Front Vowel Shift by black South Africans is an important area to expand upon in this description of SAfE.

The research I report on in this thesis was thus undertaken with the aim of answering the following questions:

1. To what extent is the Reverse Short Front Vowel Shift present in SAfE?
 - a. How far has the change progressed?
 - b. Who participates in or leads the change?
 - c. To what extent do black speakers participate in this newer vowel shift, and thus what claims can be made about convergence and deracialisation?
2. Are /æ/ and /ʌ/ in a process of vowel merger?
 - a. If so, how far along is the change; if not, what leads to the overlap?
3. What patterns can be found in the changing vowels of SAfE, and to what extent do they overlap with patterns found in other Englishes?
 - a. If there are global patterns, can these be attributed to internal or external factors?

In order to answer these questions it is necessary to discuss and illustrate vowel chain shifts and mergers, with a prefatory note on relevant terminology.

1.2 TERMINOLOGY AND LABELLING

The apartheid era in South Africa has resulted in sharply defined varieties of SAfE (Mesthrie 2010, 5), which have traditionally been described as consisting of five main varieties (each with internal variation). Descriptions of these thus necessarily involve ethnic labels in their names and identifications:

- White South African English
- Black South African English
- Indian South African English
- Coloured South African English
- Afrikaans English

Da Silva (2007, 2) and Bekker (2009, 68) argue that these classifications need revision, since they are no longer accurate in describing English in post-apartheid South Africa. SAfE (the umbrella term for Englishes of South Africa) is in fact undergoing deracialisation (Mesthrie

2010). At this stage in the development of SAfE ethnic labels are still necessary to accurately assess the changing face of English in the South African context. I therefore refer to the varieties of SAfE by their traditional labels (as indicated above) in citing the literature, particularly in Chapter 2. For the purpose of the analysis in Chapters 4 and 5 it is more appropriate to speak of ‘black speakers of SAfE’ and ‘white speakers of SAfE’ rather than Black South African English or White South African English.

In reference to vowels I will use the International Phonetic Alphabet (i.e. /ɪ/, /e/, /æ/ and /ʌ/) in addition to the labels devised by Wells (1982). The vowels I am concerned with are thus labelled KIT, DRESS, TRAP and STRUT. These labels refer to the lexical sets associated with each vowel, and form a useful mnemonic and notational device for sociophonetic description. For example, TRAP refers to all words where the /æ/ vowel is realised, such as *cat*, *maths* and *badge*. As a result of the analytical tool used (FAVE), vowels are also referred to by their ARPAbet transcriptions in the graphs presented in Chapters 4 and 5.⁴ Examples include AE for TRAP and IH for KIT.

1.3 THE CHAIN SHIFT AND MERGER OF VOWELS

Vowel chain shifts and mergers are often discussed together since they are dichotomous outcomes of a particular type of language change (Gordon 2002, 244). One cannot therefore consider vowel chain shifts in a language without considering the possibility of vowel mergers, and vice versa. Both processes involve the movement of phonemes out of their original spaces into or towards the spaces of other phonemes. Vowel merger occurs when the movement of one phoneme into the ‘space’ of another causes a loss of distinction between the two phonemes. If the movement of one phoneme into the space of another causes it to move to maintain phonemic distinction, a chain shift has occurred. Both mergers and chain shifts affect vowels that occur in the same phonological neighbourhood, such as short front vowels or low short vowels (Gordon 2002, 253).

Chain shifts also apply to consonants, a well-known example being Grimm’s Law. This law governs the changes to certain Proto Indo-European consonants in what became Proto Germanic (itself later giving rise to languages like Gothic and Old English), and is

⁴ A full discussion around FAVE and the ARPAbet is provided in Chapter 3, section 3.4.

exemplified in table 1.2, based on Hock (1991, 37). Where two sounds are provided it indicates allophonic variation.

Proto Indo-European	Proto Germanic
p, t, k	f, θ, x/h
(b), d, g	(p), t, k
b ^h , d ^h , g ^h	b/β, d/ð, g/ɣ

Table 1.2: Grimm's Law, as presented by Hock (1991, 37).

Essentially, Grimm's Law involves the spirantisation of the voiceless plosives /p, t, k/; the devoicing of voiced plosives /b, d, g/ and the allophonic deaspiration/spirantisation of /b^h, d^h, g^h/. It is unconditioned in that there are no phonological constraints on the changes involved (Hock 1991, 35). A later, also well-known, consonant shift occurred in High German, where the Proto-Germanic voiceless plosives *p *t *k shifted and became the affricates *pf, *ts, *kx, after which they potentially shift to geminate fricatives *ff, *ss, *xx (Schrijver 2011, 218).

The discussion below is however limited to vowel chain shifts as the sole focus of this thesis. It begins with a description of vowel chain shifts and mergers (section 1.3.1), followed by a discussion of the underlying principles associated with how changes occur and pattern (section 1.3.2). Thereafter studies pertaining to mergers and chain shifts in global Englishes are outlined (section 1.4).

1.3.1 DEFINITIONS AND APPROACHES

Pioneers in the study of vowel chain shifts include the historical linguists Jespersen (1949), Luick (1914-1929), Martinet (1952) and Kuryłowicz (1965). In particular, these scholars were responsible for the descriptions and analyses of the famous Great Vowel Shift, occurring in England during the time from circa 1400 to circa 1800. This vowel shift involved the change of long high vowels to diphthongs, and the subsequent raising of the remaining long vowels (see Lehmann 1962, 151-3; Hock 1991, 146-7). Otto Jespersen was the first historical linguist to connect the changes evident in Old and Early Modern English, implying causation and internal connectedness by naming it the 'Great Vowel Shift' in his 1949 work (Stockwell & Minkova 1988, 366; Labov 1994, 145). The neogrammarian Karl Luick interpreted certain phenomena such as the Great Vowel Shift as systematic (Fisiak 1988, 16). These and the interpretations of other historical scholars, though highly influential,

have been subject to much debate (see for example Stockwell & Mikova 1988; Lass 1988; Labov 1994).

André Martinet (1952) was a pioneer in studying vowel chain shifts, and was of the opinion that vowels move simultaneously in a chain-like shift to maintain distinctions (Labov 1994, 117). Martinet was particularly interested in the mechanisms of chain shifting, and was responsible for recognising that back vowels front, now a general principle of chain shifting, in addition to the principle of maximal dispersion between shifting vowels (Labov 2010, 93, 103). In reaching maximal dispersion between vowels, there are either push or pull chain shifts. Figure 1.2 is useful in visualising the potential push and pull effects of vowel chain shifts. In a pull chain, vowel C moves and becomes C'. The space vacated by this change prompts vowel B to move in order to fill it. Hence, vowel B is pulled into vowel C's vacated position and becomes B'. Vowel A is in turn pulled by the movement of vowel B and becomes A'. In a push chain, vowel A moves first, becoming A', which pushes vowel B away to avoid merger, which in turn pushes C away. In this sense, the term 'vowel chain shift' denotes changes to vowels that are linked; each change is related to and caused by another change (Hock 1991, 156).

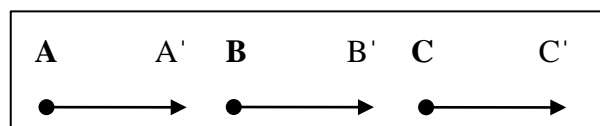


Figure 1.2 Chain shifting illustrated, based on Gordon (2011, 784).

There has been much debate over the actuality of vowel chain shifts, particularly push chain shifts. Gordon (2002, 253), for example, points to Martinet's (1952, 136) functionalist notion of maximal dispersion as an approach that has received some criticism. In particular, the arguments against push chains, and indeed chain shifts in general, would hold that the argument for the avoidance of merger in favour of chain shifting is weakened due to historical evidence suggesting that mergers are generally far more common than chain shifts (Gordon 2002, 253). Labov (2010, 140) notes that though drag chains are (now) generally accepted as possible, push chains are still "a matter of controversy." He contends that both push and pull chains are possible by citing evidence that these phenomena exist. Labov cites, *inter alia*, the Northern Cities Shift (outlined in section 1.3.2) as evidence for the existence of push chains. In so doing, he refers to the data presented in the Atlas of North American

English (Labov, Ash & Boberg 2006), which includes 62 speakers from an area exhibiting the Northern Cities Shift. The ages of the speakers range from 14 to 78, and Labov refers to a comparison of the mean F2 difference between /e/ and /ʌ/ for four age groups. The findings show that the older the speaker, the further the two vowels are from one another. Labov (2010, 141) notes that “this is strong evidence of a push chain operating within the larger mechanism of the [Northern Cities Shift].” If a pull chain were in effect, the older speakers would have the smallest distance in F2 between the vowels instead of the largest. Having established the both push and pull chains are possible, Labov (2010, 144) notes that pull chains are more frequent, as push chains are more complex.

In proving that a chain shift is in progress, two criteria must be met (Gordon 2002, 253-4): first, a distinction between the sounds involved in the movement must be maintained. Preserving contrasts is an essential aspect of both the definition and the outcome of chain shifts. Indeed, “the maintenance of distinctions is not an inadvertent consequence of chain shifting, but rather an integral dimension of the process” (Gordon 2011, 791). Thus in proving the existence of an alleged chain shift, evidence of the adjustment of vowels in response to the movement of others must be found.

As figure 1.3 illustrates, vowels can shift along multiple trajectories. These trajectories often overlap, blurring the relationship between moving vowels and the distinctions they must maintain when shifting in a chain. There are various ways of approaching this overlap in order to provide the necessary evidence of chain shifting. Gordon (2002, 256) suggests that phonemes in a chain shift such as the Northern Cities Shift (figure 1.3) can have overlapping allophonic distributions.

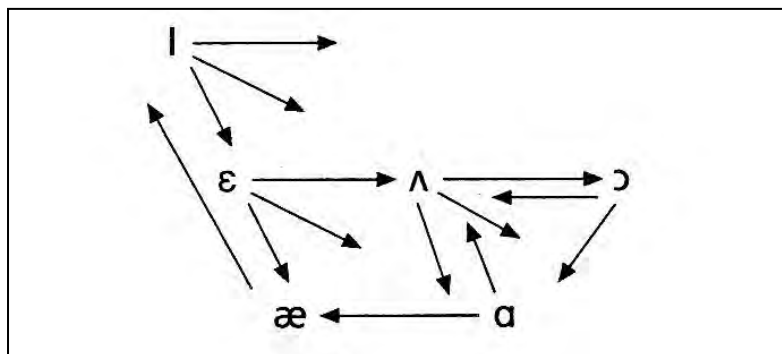


Figure 1.3: Overlapping trajectories in the Northern Cities Shift (Gordon 2002, 255).

Consequently, despite considerable overlap between the trajectories of shifting vowels, allophonic distribution preserves the distinctions between the shifting vowels: while there are allophonic realisations of the shifting vowels that overlap, there remain allophonic realisations that do not. In situations like these it is thus imperative to determine whether the shift operates on a phonemic or an allophonic level (Gordon 2011, 792). The question is therefore whether a whole phoneme or only certain of its allophones are involved in the chain shift.

The second criterion in proving the existence of a vowel chain shift originates from the assumption that there is a causal relationship between each sound involved in a chain shift, and requires these sounds to be interrelated. Interrelatedness between sounds comprises connections within the following dimensions: (a) temporal, (b) spatial and (c) the speech of the individual (Gordon 2002, 257-64).

Temporal relations between sounds in a potential chain shift involve relatedness in time. Proving the chronological ordering of changes is essential in determining the causal relationships between changing vowels (Gordon 2002, 257). Therefore, to prove the occurrence of the chain shift supposed in figure 1.1 is to prove that, for instance, B changed first which subsequently led to the changes in A and C. The actuation of a change is notoriously difficult to pinpoint, especially if the sample studied is not comprised of real time data.

In addition to relatedness in time, vowels involved in a potential chain shift must be related spatially: evidence of time relations or sequential ordering alone cannot prove relatedness between changes. Vowels in a supposed chain shift can be related spatially in geographic or linguistic terms. Geographical relatedness requires that the changes thought to be part of a chain shift be found in the same dialect, regionally or socially. It follows that if a chain shift were indeed in progress, the changes concerned would occur within the same dialect or sociolect (Gordon 2002, 260).

Linguistic relatedness between sounds involves phonetic or phonological features that are shared by the sounds in question (Gordon 2002, 260). Chain shifts would most likely occur in the vowels that belong to the same subcategory:

When one vowel shifts into the territory of another, the threat to the phonemic contrast between the vowels is greater when they represent the same subclass because this entails that they share more phonetic and phonological common ground than do vowels of differing subclasses (Gordon 2011, 789).

Vowel length, frontness and height are crucial aspects of the phonology that keeps vowel categories separate. For example, high front FLEECE is distinguished from high front KIT by length. The short front vowels KIT, DRESS and TRAP are distinguished largely by height. If the movement of one of these short front vowels compromises this distinction by encroaching on the height of another short front vowel, then a chain shift is likely to occur. In addition to the sub-categories of vowels, phonological conditioning is an important aspect to consider. Allophonic distributions are relevant during chain shifts if some environments favour a particular change while others do not (Gordon 2011, 792).

Since vowels involved in a chain shift are related in terms of time and space, it is to be expected that the changes observed and posited as part of a chain shift be present in the speech of the individual. Changes that are related to one another would likely co-occur in an individual's speech and also across the speech of individuals with relative frequency. The use of one change and not of another within an individual's speech would be unexpected, as would inconsistent use of changes across speakers (Gordon 2002, 261). If Vowel B changes and pushes Vowel C away, both changes would be present within an individual's speech as well as across individuals' speech, and their use would pattern similarly. Though a focus on the role of the individual in chain shifting is subject to debate, Gordon (2002, 261) mitigates the importance of considering the individual based on the work of Martinet (1952) and Labov (1994).

Where vowel mergers are concerned, Strelluf (2014, 20) notes that they are usually described as occurring in three patterns: merger by transfer, merger by approximation and merger by expansion. Labov (1994, 321-3) discusses these patterns with the aim of defining the various routes via which two phonemes merge. Labov (1994, 323) notes that "merger by transfer is

the slowest; merger by approximation may take three or four generations; merger by expansion appears to be complete in a single generation.”

When vowel merger occurs by transfer, the phonemic categories of words change individually in a unidirectional way, resulting in the tokens of one vowel class ‘abandoning’ their position in favour of another vowel’s position. In this sense, words are “transferred gradually from one phonemic category to another” (Labov 1994, 321). He cites the merger of /a/ and /ɑ/ in Belfast English (Milroy 1980) as an example of a merger spreading word by word.

Merger by approximation is a process most similar to chain shift movements. In this kind of merger, two vowel classes move towards each other and become indistinguishable from one another. There are various ways in which merger by approximation can manifest itself (Labov 1994). In French, /a/ and /ɑ/ combined (Lennig 1978), resulting in a single vowel positioned between the two original positions of the merging vowels. In New York City, however, the merger of /ɔɪ/ and /ʊɪ/ did not show such intermediate positioning: “the merged phoneme has the same mean value as one of the members of the merger, but with an enlarged class membership” (Labov 1994, 321).

Merger by expansion occurs when the distinction between two vowels is lost, but neither vowel loses its original position in the vowel space. The result is a single vowel allophonically distributed across the range previously divided between two separate phonemes. As an example, Labov (1994, 322) cites the work done by Herold (1990) on the merger of /a/ and /ɔ/ in Tamaqua speakers from Eastern Pennsylvania: “all merged speakers produce tokens of the new phoneme that are [ɑ]-like, tokens that are [ɔ]-like, and tokens that are intermediate in quality.”

Where vowel mergers have traditionally been seen as structural phenomena, Gordon (2002) outlines numerous reasons for varied approaches to their investigation. He argues that it is important to study the role of both production and perception in mergers, since the loss of distinction between two vowels involves the ability by the speaker to both perceive and produce the merger. In contrast to historical linguistics, the variationist approach to the study of vowel merger has brought to light new understanding of the operation of change in progress in terms of the linguistic system and the perspective of the speaker (Gordon 2002,

246). The two indicators of vowel merger are therefore (a) how speakers categorise sounds and (b) how speakers produce those sounds (Gordon 2002, 250).

Vowels can merge conditionally or unconditionally. In unconditional mergers, where a phonemic contrast between two sounds is lost in all environments, a speaker would produce and hear the merged sound in all contexts. A widespread unconditioned merger in American English is Low Back Merger, where /ɑ/ and /ɔ/ merge in all environments (Labov 1994, 316ff). In other words, a merged speaker would make no distinction between COT and CAUGHT, and would not hear a distinction in the speech of a merged speaker. Conditional mergers, where the merger is constrained to specific phonological environments, would see a speaker producing and hearing the merged sound contextually and the unmerged sounds otherwise (Gordon 2002, 245). An example of a conditioned merger is the merging of /ɪ/ and /ɛ/ pre-nasally in some varieties of American English (Labov, Ash & Boberg 2006, 65). Speakers who participate in this merger would therefore produce and perceive PIN and PEN as /ɪ/, though they would use and hear /i/ for KIT and /ɛ/ for DRESS.

However, some speakers identify words as being distinct in a minimal pairs test though they produce them as merged; and yet others produce them as distinct while identifying them as merged. The fact that speakers produce a distinction but hear a merger has left linguists confused, though much evidence in support of these near or apparent mergers has and continues to surface (Gordon 2002, 249). The first reported instance of this phenomenon, called near-merger, is illustrated by a speaker called Dan Jones. Labov, Yaeger and Steiner (1972) investigated Jones, among others, with regards to the supposed merger of /ul/ and /ɔl/. Jones, a teenager from New Mexico, was given a minimal pairs test and was asked to read words aloud to be used in a commutation test. He deemed words such as *fool* and *full* the same. However, when his realisations were acoustically analysed there were small but consistent differences in F2 between the vowels, showing that Jones produces a distinction that he does not hear (Labov 1994, 360-362).

Di Paolo and Faber (1990) studied the near-merger of certain long and short vowel pairs before /l/ (/i ~ ɪ/ ; /u ~ ʊ/ ; /e ~ ɛ/) in Utah English. They examined the acoustic properties of these near-mergers, and found that, though an F1/F2 distinction between the vowels is lost, a merger is avoided by way of phonation: breathiness or creakiness distinguishes one vowel from the other. This study, among others, provides evidence of the existence of near-merger

on a large scale within a community. In addition, it emphasises the importance of considering factors beyond F1 and F2 when studying mergers (Gordon 2002, 250).

Labov (1994, 295ff) discusses near-mergers with regards to an “impossible unmerging” in Early Modern English, where the apparent merger of *meat* (/ɛ:/) and *mate* (/a:/) seemed to have been undone by the subsequent merging of *meat* with *meet* (/e:/). The undoing of mergers is impossible according to Garde’s Principle, which states that mergers are irreversible by linguistic means (Strelluf 2014, 20).⁵ Milroy and Harris (1980) studied this phenomenon in Belfast English, and based on their findings, in conjunction to his and others’ work on Early Modern English, Labov (1994, 387) concludes:

The reports of merger dating from 1700 in Belfast then are actually reports of near-mergers, and imply that the London reports of the 16th century concerned near-mergers as well. These indicate an asymmetry of production and perception that is common to both communities (...).

The notion of near-merger accounts for the unexpected instance of a speaker producing a difference that he or she does not perceive. Importantly, it has emphasised the need for an expanded understanding of the role of perception in mergers, which must now be seen as including the subconscious (Gordon 2002, 251).

The reasons behind this disconnect between perception and production should be investigated, according to Gordon (2002, 251). He refers to Milroy’s (1992) study of Hiberno-English as a possible answer. Milroy claims that speakers of Hiberno-English have access to two phonological systems and are able to choose between them. Gordon (2002, 252) concludes that near-merger should be viewed as a situation where speakers accept an incoming norm in stages: they accept the merger perceptually, and the production thereof comes at a later stage. Labov (1994, 363) provides support for this notion via a speaker called Bill Peters who style shifts between merged and unmerged realisations of COT (/ɑ/) and CAUGHT (/ɔ/). Peters uses more merged realisations in the formal minimal pairs test and unmerged versions in spontaneous speech. Peters was 80 years old at the time of Labov’s study, suggesting that the second step in complete merger (production matching perception)

⁵ The full set of principles governing vowel chain shifts and mergers are provided and discussed in section 1.3.2.

may never occur, which in turn shows that near-mergers “need not be seen as transitional situations preceding complete mergers” (Gordon 2002, 252).

With regards to the Reverse Short Front Vowel Shift and the potential merger of TRAP and STRUT in SAfE, there are various boxes to be ticked in proving their existence and elucidating their exact nature. Labov (1994) has formulated principles and patterns that govern chain shifts and mergers, shedding light on possibilities and patterns to be observed.

1.3.2 LABOV’S PRINCIPLES AND PATTERNS FOR VOWEL SHIFT AND MERGER

Labov (1991, 3) attributes the diversity in modern vowel systems (including English and German) to three phonetic changes: chain shifts, mergers and syllabic shifts. Each of these is constrained by certain principles, those pertaining to shifts and mergers being my main concern. These principles are set out by Labov (1994) and are based on empirical evidence (Labov & Wald 1969; Labov et al. 1972). They are by no means universal: they are principles that aid us in predictions of possible changes as well as estimations of the likelihood of changes occurring in certain ways (Labov 1991, 34-5).

A key tenet of Labov’s principles is the Peripherality Hypothesis, an acoustic measure that places vowels along one of two tracks in the vowel space: an inner, non-peripheral track and an outer, peripheral track (Labov 1991, 6), as is seen in figure 1.4.

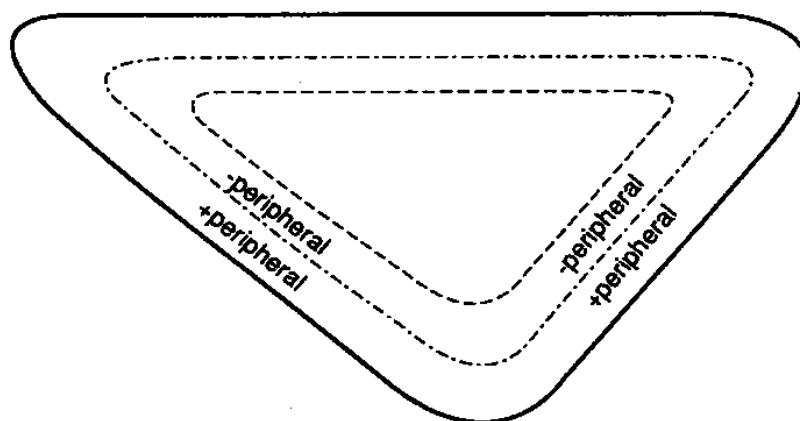


Figure 1.4: The [\pm peripheral] tracks followed by changing vowels (Labov 1994, 177).

Usually, tense/long vowels are [+peripheral] and lax/short vowels are [-peripheral] (Thomas 2011, 280). Labov et al. (2006, 17) state that the Peripherality Hypothesis has two propositions: (1) English vowel spaces have [\pm peripheral] areas at the front and at the back; and (2) in chain shifting, vowels rise in the peripheral areas and lower in the non-peripheral areas (as illustrated in figure 1.5).

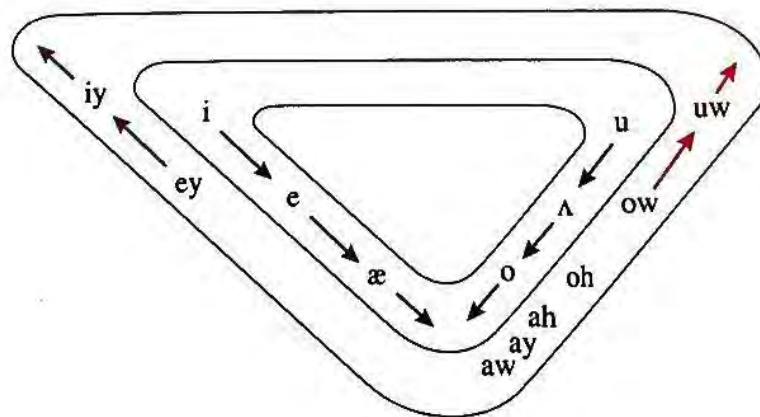


Figure 1.5: Typical directions of vowel shifts along the [\pm peripheral] track. (Labov, Ash & Boberg 2006, 16).

Labov (1991, 6) posits peripherality and the acoustic space as the basis for the principles of chain shifting. He argues that distances between vowels are maintained resulting in distinct high, mid, low, front, central and back vowels despite numerous sound changes, showing that there must be principles in operation across vowel subsystems. Most chain shifts involve changes in height, frontness and rounding within a single subsystem (Labov 1991, 7-8).

The discussion surrounding the principles in Labov (1994, 116) starts with three general principles indicated and predicted by Sweet (1888, 19-21). Haudricourt and Juilland (1949) and Martinet (1955) reintroduced these principles to the study of vowel chain shifts, and they are currently applicable to many vowel changes in progress (Labov 1994, 117). To these three principles (I-III in table 1.3) are added five more on chain shifting and two on mergers, following various revisions by Labov and Wald (1969) and Labov (1991, 1994).

Principle	Application
I	Tense nuclei rise along a peripheral track.
II	Lax nuclei fall along a non-peripheral track.
IIa	The nuclei of upgliding diphthongs fall.
III	Tense vowels front along peripheral paths, and lax vowels back along non-peripheral paths.
IV	Low non-peripheral vowels become peripheral. (Lower Exit Principle, Labov et al. 2006, 18).
V	One of two high peripheral moras becomes non-peripheral. (Upper Exit Principle, Labov et al. 2006, 18).
VI	Peripheral vowels rising from mid to high position develop inglides.
VII	Peripherality is defined relative to the vowel system as a whole.
VIII	Elements of the marked system become unmarked.
Garde's Principle	Mergers are irreversible by linguistic means.
Herzog's Principle	Mergers expand at the expense of distinctions.

Table 1.3: Principles governing chain shift and merger.

Based on chain shifts in progress in various languages, these principles are combined as Chain Shift Patterns (Thomas 2011, 282), as shown in table 1.4. Labov (1994) considered various English dialects in the application of these principles, thus characterising various shifts in progress. These are represented in figures 1.6-1.8 and my descriptions rely largely on Thomas (2011, 282-3). Roman numerals indicate which of Labov's principles are in application with each shifting vowel.

Pattern	Principles	Application
1	I, II	Long, high vowel diphthongises and another vowel rises to take its place.
2	I, II, III	Fronting and raising of long vowels with falling short vowels.
3	I, III	Fronting back vowels with raising vowels to take their place.
4	-	Reversal of peripherality.

Table 1.4: Labov's chain shift patterns.

The Northern Cities Shift (figure 1.6) occurs in the Great Lakes area of the United States of America, and involves, among other things, KIT centralisation and lowering (under principles II and III); DRESS lowering and backing (under principles II and III), TRAP raising (under principle I) and STRUT backing (under principle III).

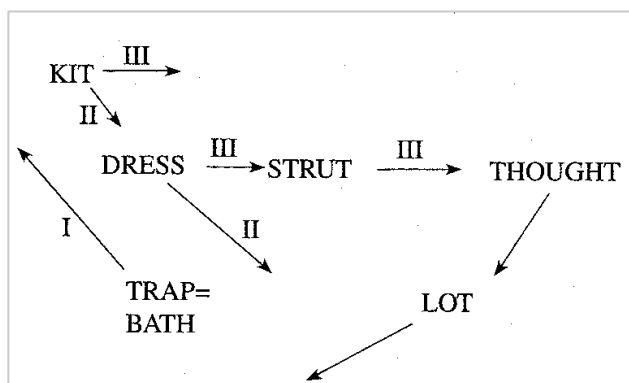


Figure 1.6: The Northern Cities Shift as illustrated by Thomas (2011, 282).

The Southern Shift (figure 1.7) involves, among other changes, KIT and DRESS raising (under principle I), GOOSE fronting (under principle III) and FLEECE lowering (under principle V and IIA). According to Labov (1994, 202), this shift is in operation in southern parts of the United States, the Southern Hemisphere (including South Africa) and southern England.

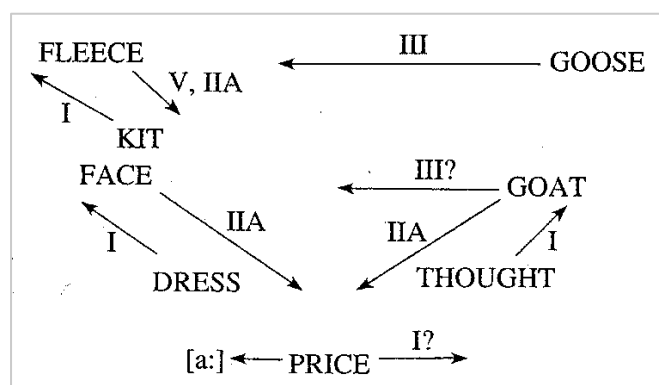


Figure 1.7: The Southern Shift as illustrated by Thomas (2011, 282).

The Canadian Shift (figure 1.8) is present in the English of Canada and certain Midwestern parts of the United States of America, and involves KIT, DRESS and TRAP lowering (all under principle II) Labov (2010, 16) notes that this shift also involves DRESS backing, and may involve KIT backing (under principle III). Regarding the Reverse Short Front Vowel Shift in SAfE, the features of the Canadian Shift appears to be more suited to current SAfE than the Southern Shift, which is more similar to the Older Short Front Vowel Shift.

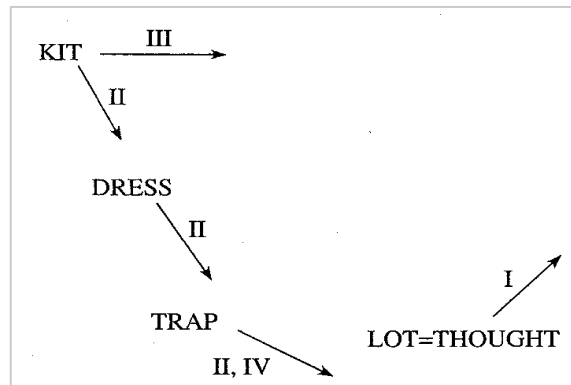


Figure 1.8: The Canadian Shift (Thomas 2011, 283).

Various scholars have called the validity of these principles into question (such as Cox 1999 and Thomas 2003). Torgersen and Kerswill (2004) challenge the principles in that they are internal to a language without taking external factors into account. Though they acknowledge that Labov’s principles comprise “an explicit framework for the systematic exploration and comparison of chain shifts” they maintain that external factors have the ability to trump natural or internal processes (Torgersen & Kerswill 2004, 28; 47). Bekker (2009, 55) approaches the patterns and principles as a working model to use “in order to capture undeniable regularities and similarities across different languages and accents.” I thus utilise the principles and patterns as guide in two ways: first, with regards the Reverse Short Front Vowel Shift, the framework allows for a certain degree of prediction based on precedent: while there is no impossible direction of change, some directions are more frequent than others (Labov 1994, 116). Second, I use the framework to aid the comparison of SAfE to patterns found in other Englishes.

1.4 GLOBAL VOWEL CHANGE PATTERNS

To a large extent, the British spread English throughout the world via colonisation starting in the 1600s. In this way English was transported from its land of origin to various parts of the world where it took shape as mother tongue Englishes referred to as Extraterritorial Englishes (Lass 2004, 363). As a result of the history of colonisation, Extraterritorial Englishes (ETEs) are comprised of two groups, namely Northern Hemisphere ETEs and Southern Hemisphere ETEs (Lass 2004, 369). Englishes from America, Canada and Ireland are Northern Hemisphere ETEs, and the Southern Hemisphere ETEs include Englishes from South Africa, Zimbabwe, Australia and New Zealand. According to Lass (2004, 370), the southern ETEs are largely ‘British’, both phonologically and lexically, compared to Northern Hemisphere ETEs, which are more ‘American’. Phonological examples of this include the pronunciation of *beta* as /bi:tə/ rather than /beitə/, as well as the retention of /h/ in the word *herb*. Lexical examples include *petrol* for *gas*, and *bonnet* for *hood*. America largely cut itself off from the community in Britain in 1776 (during the American Revolution), which Lass posits as the reason for this difference. Countries in the southern hemisphere have maintained close ties with Britain throughout most of the 1900s, and he notes that one can “find out more about the Royals from the *Cape Times* than from the *Washington Post*” (Lass 2004, 370).

The third aim of this research (section 1.1) is to situate SAfE globally; that is to describe the linguistic changes evident in SAfE and to compare them to patterns observed in ETEs and British Englishes. In comparing the similarities between SAfE and other Englishes, I aim to show whether the linguistic changes observed are internally driven as explained by Labov’s principles, or externally driven, along the same lines as argued for by Torgersen and Kerswill (2004). The following discussion thus details linguistic descriptions of English in the United Kingdom in section 1.4.1 and ETEs from the Southern Hemisphere (New Zealand and Australia) in section 1.4.2. In each case I will note whether these varieties can still be considered part of the Southern Shift as posited by Labov (2004, 202). In section 1.4.3 I discuss the Northern Hemisphere ETEs (Ireland, the USA and Canada), paying particular attention to further developments to the chain shifts thus far identified for North America and Canada.

1.4.1 THE UNITED KINGDOM

In introducing their study of short vowels, Torgersen and Kerswill (2004) outline various studies providing evidence for the lowering of the short front vowels in Southern British English (e.g. Wells 1982, Bauer 1994, Trudgill 1999, 2004). This lowering is posited as part of a pull chain initiated by the lowering of TRAP (Trudgill 2004). To further investigate the chain shift, they studied the vowels of speakers from two areas in southeast England: Ashford and Reading. Their data was comprised of two age groups, 14-15 and 70-90, and was almost evenly split between the genders (Torgersen & Kerswill 2004, 34-36).

In Ashford, they discovered ‘anti-clockwise’ vowel movement where FOOT /ʊ/ is fronted, DRESS and TRAP are lowered, and STRUT is backed and raised. The beginnings of this vowel shift are evident in the older speakers who exhibit slight FOOT fronting, central STRUT and slight lowering of DRESS and TRAP. The younger speakers (presented in figure 1.9) have taken the shift a step further with even fronter FOOT, backer STRUT, raised LOT and lower DRESS and TRAP realisations (Torgersen & Kerswill 2004, 41). In conjunction with the findings of other studies, these results confirm that this kind of vowel change, i.e. the Southeast English Chain Shift, is common in southeast England. They suggest that the chain shift was initiated by lowering TRAP, which pulls DRESS along and pushes STRUT away (Torgersen & Kerswill 2004, 45).⁶

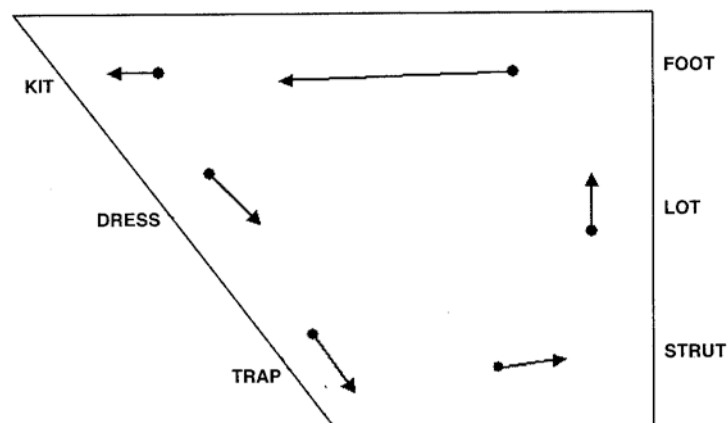


Figure 1.9: Chain Shift in Ashford, United Kingdom (Torgersen & Kerswill 2004, 40).

⁶ These changes to TRAP and STRUT are also reported for Received Pronunciation (RP) by Fabricius (2007).

In Reading, however, the results are different, as is evident in figure 1.10. There is no evidence of a chain shift in the short vowel system, although STRUT and FOOT lower and front respectively.

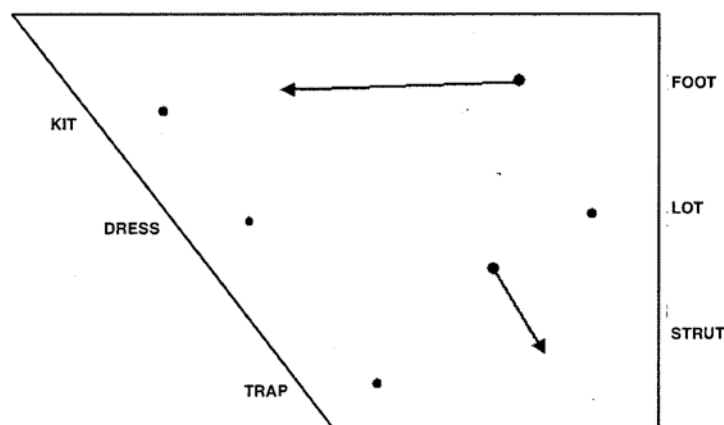


Figure 1.10: Vowel changes in Reading, United Kingdom (Torgersen & Kerswill 2004, 45).

Though the changes to the Reading system do not mirror the chain shift evidenced in Ashford, Torgersen and Kerswill (2004, 44) attribute this difference to dialect convergence, given that the changes in Ashford and Reading result in vowel systems that are very similar. Many Londoners have emigrated to Ashford and Reading, thus exposing both places to the norms of English in London. The convergence required the speakers in Ashford to participate in the Southeast English Chain Shift, whilst the only changes necessary in Reading were FOOT fronting and STRUT lowering.

In 2006 Torgersen, Kerswill and Fox expanded the study to include London speech as a way of testing the 2004 findings. They consider linguistic data from three pre-existing datasets which gave rise to information from the following areas in London: Tower Hamlets and Hackney (east London); Camden and Barnet (north London) and south and west London.⁷ In addition to these, they interviewed speakers from Hackney and Havering, representing inner and outer London respectively.

⁷ These datasets are: Intonation Variation in English; Corpus of London Teenage Language and Labov's 1968 London recordings (Torgersen et al. 2006, 5).

Their study confirms not only that the vowel shift apparent in Ashford is present in London, but also confirms Trudgill's (2004) postulation of the Southeast England Chain Shift (Torgersen et al 2006, 11). There are various degrees of participation in the changes by speakers per area, which the authors attribute to a gradual spread of the change: it is yet to be fully taken up in certain areas (Torgersen et al. 2006, 12). They attribute surprising characteristics (such as the lack of FOOT fronting in Hackney despite extreme GOOSE fronting) to language contact, particularly with non-native and ethnic varieties of English in addition to other sociolinguistic factors (Torgersen et al. 2006, 12-13). Further support for the anti-clockwise short front vowel shift in southeast England is provided by Kamata (2008), who finds that this shift is well advanced in the speech of young, upper middle class London females.

Though this chain shift does not contradict Labov's principles, particularly II and III, it runs counter to the suggestion that southeast England participates in the Southern Shift. In this sense, the vowel changes to southern British English are better characterised as equivalent to the Canadian Shift.

1.4.2 THE SOUTHERN HEMISPHERE

Varieties of English in South Africa, New Zealand and Australia are said to exhibit features of the Southern Shift along with Englishes in southern England and parts of the southern United States. Though these dialects have been traditionally viewed as quite different from one another, they are all said to participate in similar kinds of vowel chain shifting (Labov 1991, 22). In addition to this commonality, New Zealand, Australia and South Africa form the cohort of Southern Hemisphere ETEs as described earlier. The South African Chain Shift (Lass & Wright 1986) is not discussed here, in favour of presenting the phonetic description of SAfE holistically in Chapter 2. In short, it involves the raising of KIT, DRESS and TRAP when compared to other varieties of English, particularly RP, and in this way conforms to the trends reported for the Southern Shift. The discussion to follow therefore starts with vowel trends in New Zealand followed by details of Australian vowel shifts.

The New Zealand English Short Front Vowel Shift (henceforth New Zealand Shift) has been well documented, and involves KIT centralisation and DRESS and TRAP raising (Watson, Maclagan & Harrington 2000, 62). There are various theories pertaining to the direction and causation of this chain shift. Watson et al. (2000, 65) argue that it is a pull chain initiated by the retraction of KIT, resulting from the crowding of the mid-high front vowel space. They suggest this in contrast to Bauer (1979) and Woods (1997, 2000) who postulate a push chain initiated by STRUT fronting. More recently, Langstrof (2006, 162) refutes this with historical evidence of TRAP raising before KIT backing, and so argues for a push chain initiated by raising TRAP. This is confirmed in Maclagan and Hay (2007, 2) in their review of existing literature, particularly of Gordon et al. (2004). General consensus reliant on historical evidence thus appears to favour a push chain initiated by TRAP raising (figure 1.11).

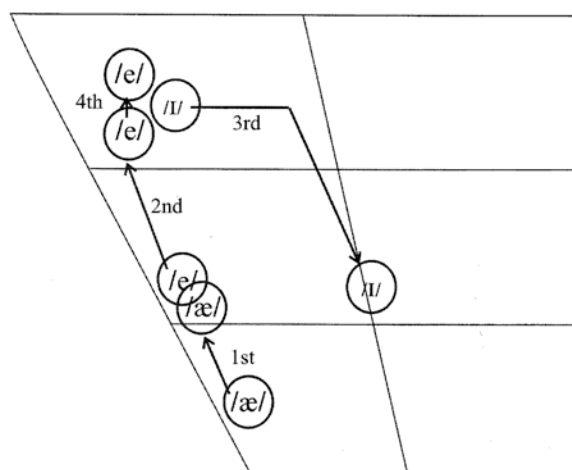


Figure 1.11: The New Zealand English Short Front Vowel Shift (Hay et al. 2008, 41).⁸

Interestingly, the New Zealand Shift also affects FLEECE (/i:/). In traditional chain shifts long and short vowels are not expected to change together, since they are not members of the same subsystem. New Zealand English shows nucleus lengthening before voiced codas, and shortening before voiceless codas, leading to DRESS in voiced environments being longer than FLEECE in voiceless environments (Hay, Maclagan & Gordon 2008, 42). An anecdotal example of this is provided in Maclagan and Hay (2007, 5): a young speaker, when in London, asked for ‘four tens’ from a bank, and was met with the response, “we don’t have

⁸ In reproducing this figure, the mid-central symbol /i/ has become slightly obscured by the line.

fourteens”. In response to this, speakers diphthongise FLEECE before voiceless codas to avoid potential confusion (Maclagan and Hay 2007, 18-9).

Raising DRESS also has as an effect on NEAR /iə/ and SQUARE /eə/, leading to “a merger of approximation on NEAR” (Hay, Drager & Warren 2009, 270). Gordon and Maclagan (2001) provide acoustic evidence for this merger via a study of school children over a period of fifteen years (1983-1998). In 1983, around 15% of the children produced merged qualities in their vowels, and in 1998 this figure rose to 90%, showing that the merger is well under way. In these cases, both *ear* and *air* sound like *ear* (Hay, Maclagan & Gordon 2008, 40-1).

Though not part of the chain shift, STRUT’s ‘behaviour’ is worthy of mention. In New Zealand English STRUT largely overlaps with START (/ɑ:/), which is realised as [a:]. Young women have recently been found to produce STRUT slightly, though significantly, higher than START. In all other cases the chief differentiating feature between STRUT and START is length (Warren 2006, 128-30).

The manner in which the vowels shift in New Zealand English violates Labov’s principles II and III (section 1.3.2) in that short vowels rise together and that a short vowel retracts (Langstrof 2006, 142). In this sense New Zealand English participates in the Southern Shift only with regards DRESS and TRAP raising.

Australian English has been described by various scholars, some of whom investigate the influence of New Zealand English on Australian English regarding short front vowel raising and KIT centralisation (for example Bradley 1989; Holmes & Bell 1990). Acoustic studies of vowel change in Australian English have been neglected according to Cox and Palethorpe (2001, 20), who set out to describe any vowel changes in Australian English whilst, among other things, exploring conformity to Labov’s principles of chain shifting. Their study involved diachronic analysis using Holmes and Bell’s 1960s data in addition to their own data from 1990; as well as synchronic analysis using the Australian National Database of Spoken Language (ANDOSL) (Millar, Vonwiller, Harrington & Dermody 1994). Their findings concerning monophthongs are summarised as follows (table 1.5):

Diachronic		Synchronic
<i>Vowel Change</i>	<i>Vowel Change</i>	<i>Participants concerned</i>
Raised /ɪ/	Fronted /i/	Females only
Lowered and retracted /æ/	Lowering /æ/	Males and females
Raised /ɒ/	Lowering /a/	Females only
Raised /ʊ/	Lowering and fronting /ʊ/	Females only
Raised and fronted /u/		
Fronted /ɜ/		

Table 1.5: Diachronic and synchronic vowel changes in Australian English (Cox & Palethorpe 2001, 25; 34).

The results of these studies show that Australian English vowels have undergone change, and that monophthongal changes are particularly evident in the lowering of /æ/ (Cox & Palethorpe 2001, 40).⁹ They furthermore negate the suggestion by Holmes and Bell (1989) that New Zealand English influenced Australian English in short front vowel raising, since the results point in the opposite direction for /æ/ and the onset of /eɪ/, which they also found to be lowering. Both of these, as discussed above, undergo significant raising in New Zealand English. The diachronic study shows a chain shift involving /ɒ/ and /ʊ/, and the synchronic study shows, for females, a chain shift involving /æ/ and /a/. Labov's principles of chain shifting are largely supported by these findings, particularly in lowering short /æ/, though they are violated by the female lowering of /a:/ and fronting of short /ʊ/ (Cox and Palethorpe 2001, 41).¹⁰

A subsequent study by Cox and Palethorpe (2008) track the vowel changes in Australian English over a period of 100 years. They make use of various studies and corpora in their comparisons: 1990 data (Cox 2006), ANDOSL (Millar et al. 1994), 1965 data (Bernard 1967), Australian Ancestors (Cox, Palethorpe & Tsukada 2004) and Australian Voices (Cox & Palethorpe 2004).¹¹ They supplement the use of these by recordings comprising the 2007 data (Cox & Palethorpe 2008, 343). Their comparative results, showing initial raising of KIT and DRESS, and then gradual lowering and retraction, are summarised in table 1.6.

⁹ Cox and Palethorpe (2007) found that TRAP has lowered so much in Australian English that that it “is now situated at the extreme open position for young Australians” (Cox & Palethorpe 2008, 342).

¹⁰ Though the Principles were reworked in Labov (2010) to include fronting short back vowels.

¹¹ See Cox and Palethorpe (2008, 342-3) for the comprehensive description of the corpora. I retain their labels.

Data frame	Recording date	Speaker Profile	Findings
Australian Ancestors	1960-1980	Born between 1885 and 1995	Raised /ɪ/ Raised /e/
vs. Australian Voices	1998	15 year old boys	Lowered /æ/
1965 data	1960s	15 year old boys	Raised /ɪ/
vs. 1990 data	1990	15 year olds	Lowered /e/ Lowered and retracted /æ/
1990 data	1960s	15 year old boys	Lowered /ɪ/
vs. ANDOSL & 2007 data	2000s 2005-2007	Females under 30 Females under 30	Lowered /æ/ and /e/ Retracted /æ/

Table 1.6: Vowel change in Australian English over 100 years (Cox & Palethorpe 2008).

The comparison between Australian Ancestors and Australian Voices shows that Australian English underwent short front vowel raising throughout the 1900s, and supports TRAP lowering between the 1960s and the 1990s, as presented in their 2001 study (Cox & Palethorpe 2008, 343). The 1990 data, when compared to the 1965 data (Cox 1999) shows a large degree of KIT raising and the lowering and backing of TRAP. DRESS lowering is almost significant ($p=0.056$) which indicates the emergence of a new change during this time (Cox & Palethorpe 2008, 344). The later comparisons, between 1990, ANDOSL and 2007 show the expansion of this trend with significantly lowered DRESS realisations in 2007. KIT has also been lowered, with the 1990s realisations being higher than those of ANDOSL and 2007. TRAP lowering and retraction is also more present in 2007 than in 1990 and in the ANDOSL corpus (Cox & Palethorpe 2008, 344). The lowering of TRAP is also attested for Australian English by Horvarth (2004), who notes its emergence in the 1960s and 1970s.

The third comparison (1990, ANDOSL and 2007) shows that the front vowels in Australian English are lowering where they were raising before. This constitutes a reversal of former vowel raising; a reverse vowel shift. Cox and Palethorpe (2008, 345) postulate that both the initial raising shift (push chain) and the reverse chain shift (drag chain) were initiated by changes to TRAP, which may give insight regarding the causation of the Reverse Short Front Vowel Shift in SAfE. The reversal of raising removes Australian English from the Southern Shift, placing the variety more in line with the patterns observed in the Canadian Shift (figure 1.8, section 1.3.2).¹²

¹² It would however appear that the Southern Shift itself is becoming less ‘southern’ – discussed in section 1.4.3.

1.4.3 THE NORTHERN HEMISPHERE

Hickey (1999, 2005) describes ‘the vowel shift of the 1990s’ in Dublin English, which came about as a method of dissociation from traditional or conservative views of being Irish (Hickey 2005, 47). This shift involves most long vowels and diphthongs, though his description focuses on PRICE (/aɪ/), CHOICE (/ɔɪ/), LOT (/ɒ/) and THOUGHT (/ɔ/). These changes constitute a chain shift in that several segments are affected “by a process of retraction and raising” (Hickey 1999, 274). The beginnings of the shift were found in a ‘posh’ suburb, Dublin 4, during the 1980s (Hickey 2007, 355), and involves the backing of diphthongs with low back onsets, and the raising of low back vowels:

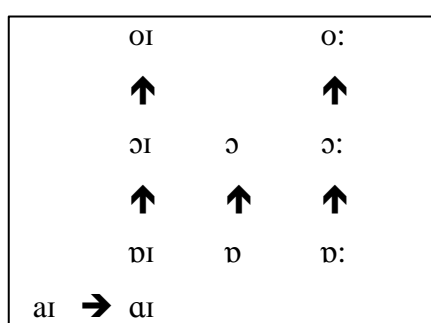


Figure 1.12: The 1990s Dublin Vowel Chain Shift (Hickey 2005, 49).

In the 1990s shift, KIT and DRESS show no differentiation from other dialects of Irish English (Hickey 2005, 50). In the late 1700s, TRAP (after velars) was a raised vowel (*gather* ~ *gether*), though throughout the 1800s this changed, with a lowered form existing by the early 1900s. He terms this kind of reversal ‘ebb and flow’ in that it can lead to backwards or forwards shifts, not merely shifts in a single direction (Hickey 2005, 71-2).

More recently, Hickey (2013) documents the lowering of the short front vowels involving mostly DRESS and TRAP, with TRAP backing before velars. KIT is also involved, though not to the same extent, as indicated by the question mark in figure 1.13. The participants in this shift are young, female non-vernacular speakers who are concerned with appearing fashionable. No substantial evidence of this shift was found in the speech of males (Hickey 2013).

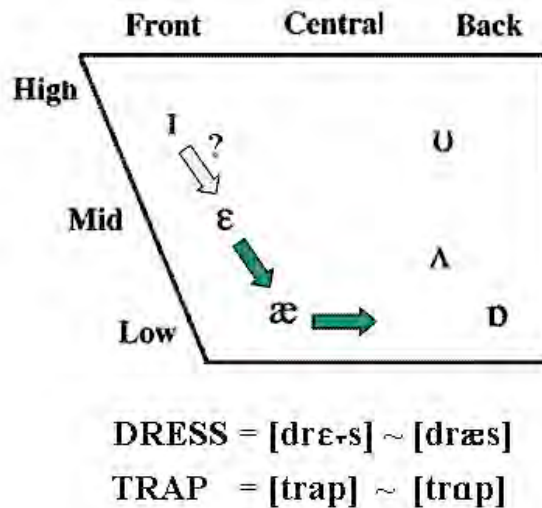


Figure 1.13: The current chain shift in Dublin English (Hickey 2013).

Together with LOT and THOUGHT raising in the 1990s and well-established GOOSE-fronting in Dublin English (Hickey 2013), a chain shift such as is described by Labov’s pattern 2 (section 1.3.2) is seemingly underway in this variety of Irish English. As is the situation in southeast England and in Australia, the patterns of vowel shifting in Dublin indicate overlap with the Canadian Shift more than the Southern Shift.

As far as the United States of America is concerned, Labov (1991) classed the varieties of American English into three broad dialect systems based on how they relate to two ‘pivot points’. The first involves the splitting of the TRAP vowel into tense and lax allophones (/bæ:d/ vs. /bæt/, for instance), and the second is Low Back Merger (Labov 1991, 12). Low Back Merger, as noted before, is an unconditioned merger between the vowel subsets LOT and THOUGHT, where words like *Don* and *Dawn*, *cot* and *caught*, and *hock* and *hawk* rhyme (Strelluf 2014, 19). Speakers with Low Back Merger have one phoneme, LOT~THOUGHT, where there were two before (see Herold 1990; Labov 1994).¹³ Labov’s first system participates in the Northern Cities Shift, necessarily demonstrating the following in terms of the pivot points: (1) the TRAP vowel is tense/long; (2) there is no Low Back Merger (Labov 1991, 14). The tensing of TRAP is what allows it to move up along the peripheral track, leaving the low front area open, causing the falling DRESS and KIT vowels (Labov 1991, 15).

¹³ Low Back Merger is also referred to in the literature as COT~CAUGHT merger or LOT~THOUGHT merger, fairly interchangeably.

The second system involves participation in the Southern Shift, described by scholars such as Feagin (1986) and Wolfram and Schilling-Estes (1998). British dialects involved in the Southern Shift demonstrate TRAP that is not split into tense and lax allophones. American dialects belonging to this system include those particularly in the Middle Atlantic States (New York, Philadelphia). These dialects split TRAP into tense and lax allophones, though lax TRAP remains stable. Tense TRAP raises in a similar way to what has been observed for the Northern Cities Shift (Labov 1991, 22). In New York City and Philadelphia, tense and lax TRAP presents a ‘true lexical split’ (Labov 1991, 22), as shown in the work of Trager (1940) and Labov (1981). Varieties belonging to the second system do not participate in Low Back Merger. According to Labov (1991, 28), despite the variety of options available to participants in the Southern Shift, /i/, /e/ and /æ/ always become peripheral (under principle VI) after which they front and raise. It is in this sense that we can conclusively say that southeast British English and certain Southern ETEs (such as Australian English and SAfE) are no longer part of the Southern Shift.

Labov’s third system is found in some parts of the United States and Canada. These varieties participate in the Canadian Shift, first described by Clarke, Elms and Youssef (1995). Here, TRAP remains largely unchanged and Low Back Merger is present (Labov 1991, 30). Since TRAP does not undergo lengthening or raising, the initiating feature of the Northern Cities Shift is not present, and due to the presence of Low Back Merger, it cannot be characterised as displaying the Southern Shift either (Labov 1991, 34). As mentioned before, varieties of English in Australia, Ireland and southeast England are reversing the Southern Shift and are now more in line with what is posited for this third system, namely the Canadian Shift.

American English varieties are also in a process of reversing these well described vowel shifts. A recent study by Dodsworth and Kohn (2012) investigates the reversal of the Southern Shift with regards the front vowels in Raleigh, North Carolina. Their findings show that the middle and upper class population are becoming less ‘southern’ in their pronunciations over a three-generation period (Dodsworth & Kohn 2012, 231). They attribute this change to large-scale language contact with non-Southern dialects of English due to immigration of workers in the field of technology (Dodsworth and Kohn 2012, 221). In 2011, Jacewicz, Fox and Salmons conducted a study that investigated vowel change in three communities – each of which were characteristically described as participants in the Northern Cities Shift, the Southern Shift and the Canadian Shift respectively. In other words, they

studied three cities, each a member of one of Labov's three dialect systems. They found changes across all three dialect areas/systems involving the overall lowering of KIT, DRESS and TRAP in the speech of adults aged between 20 and 65 (Jacewicz et al. 2011, 692). While the lowering of these vowels is characteristic of the Northern Cities Shift and the Canadian Shift, it is not for the Southern Shift, where KIT and DRESS vowels are raised. In this sense, speakers who form part of the community characterised by the Southern Shift are becoming less southern in their lowering of KIT, DRESS and TRAP.

Various scholars have described the vowel system of English in California, though Kennedy and Grama (2012, 42) aim to provide updated quantitative descriptions of the phonetics and phonology, which they view as lacking. They cite, for example, Gordon (2004), who does not provide acoustic evidence to support his qualitative description of the California Shift.¹⁴ They furthermore contend that though Eckert (2004) provides acoustic data, she does not overtly characterise the short front vowel shift via these data. They also cite Ladefoged (2006) who provides acoustic evidence for shifted vowels in the speech of Californians, but who does not make overt claims regarding region. In their study of the California Shift, Kennedy and Grama (2012, 41) use Labov et al. (2006) as a baseline for diachronic comparison.

Labov, Ash and Boberg (2006) describe a developing variety in the Western United States where Low Back Merger is a characteristic feature, and GOOSE is centralised (Kennedy & Grama 2012, 41). They identify the lowering short front vowels as indicative of the Canadian shift, thus characterising English in the Western United States as participating in this shift. However, only nine speakers in their database produce this shift, none from California (Kennedy & Grama 2012, 41). According to Labov et al. (2006, 217-218), a variety is considered as participating in the Canadian Shift when these three acoustic criteria are met:

1. The first formant (F1) of DRESS must exceed 650 Hz.
2. The second formant (F2) for TRAP must be lower than 1825 Hz.
3. The F2 of LOT~THOUGHT must not exceed 1275 Hz.

Kennedy and Grama (2012) use these parameters to characterise and analyse the speech in their sample of 13 mostly undergraduate students, native to California. They assume that speakers who do not participate in the chain shift realise the KIT vowel as [ɪ], DRESS as [ɛ], and TRAP as [æ]. Shifted speakers would have [ɛ], [æ] and [æ] or retracted [a], respectively

¹⁴ The authors reference 2006 which refers to the Multimedia Tool associated with the 2004 print publication.

(Kennedy & Grama 2012, 43). Figure 1.14 shows the results of their study. The lines drawn indicate the parameters of Labov et al. (2006) that they are comparing their data to, and for clarity, I added labels for KIT, DRESS, TRAP and LOT.¹⁵

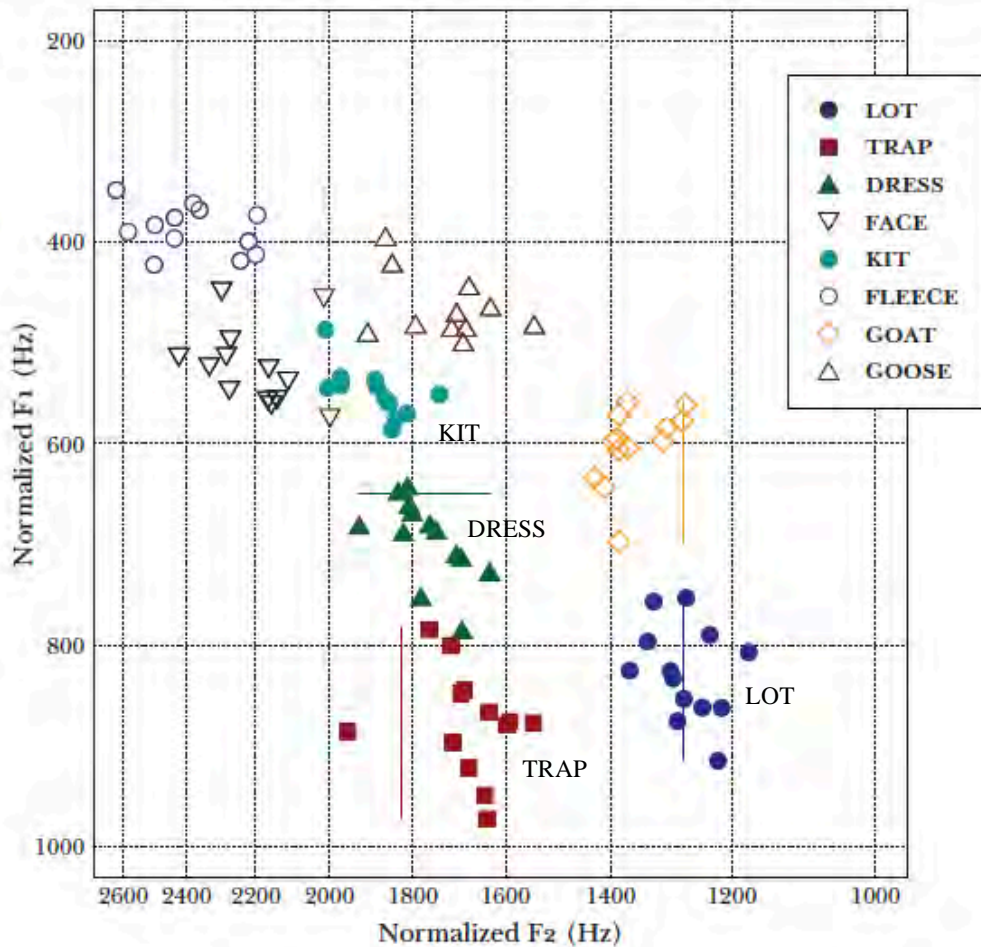


Figure 1.14: The California Shift among 13 speakers aged 19 to 29 (Kennedy & Grama 2012, 47).

Their results show that KIT and DRESS are lowering, and are realised within the parameters set for Canadian Shift, though LOT is grouped around either side of its parameter. The lowering of KIT and DRESS would suggest that Californian English as a dialect is part of the Canadian Shift, which scholars (such as Eckert 2004) believe to be initiated by Low Back Merger. The merger caused LOT~THOUGHT to retract, which in turn left space for the lowering and retraction of TRAP and subsequently DRESS and KIT. However, in Canadian English LOT~THOUGHT is much more retracted than it is in California English: the results in figure 1.13 show variation in Californian English between a central and back realisation. The

¹⁵ In their graph, 'LOT' refers to the merged value of LOT~THOUGHT.

California Shift is therefore not attributed to Low Back Merger, and is in this sense a separate entity from the Canadian Shift (see Kennedy & Grama 2012, 51).

Low Back Merger is a key aspect involved in numerous vowel chain shifts in the United States. In Kansas City, the merger between COT and CAUGHT is the potential cause of the subsequent lowering and retraction of TRAP (Strelluf 2014, 383). In other American English varieties, the link between Low Back Merger and TRAP retraction has been documented for Canada (Clarke, Elms & Youssef 1995), California (Eckert 2004) and Columbus, Ohio (Durian 2012), among others (see Strelluf 2004, 176ff). Of all the varieties of English discussed thus far, certain American and Canadian dialects are the only dialects that display Low Back Merger. Most of the Englishes previously described as participating in the Southern Shift are no longer doing so, and instead participate in something more similar to the Canadian Shift or the California Shift. Both of these chain shift patterns hinge to a certain extent on Low Back Merger, which leads one to question the instigator of these kinds of shifts in varieties that have separate COT and CAUGHT realisations. The California Shift is distinguished from the Canadian Shift in that the behaviour of COT~CAUGHT is different (though merged, more central than back), and it seems that the other varieties also behave differently. Short front vowel lowering occurs with or without Low Back Merger, though it seems that raised LOT is a common feature (certainly in southern British English, Dublin English and Australian English). In varieties without Low Back Merger, the positioning of LOT is thus is a likely contributor to short front vowel lowering.

Having discussed and described the trends in vowel shifting of the southern British English and the ETes, it is clear that the lowering of KIT, DRESS and TRAP relative to their earlier positions is a global tendency. It is also clear that changes have occurred since Labov's (1991) description of international participation in various chain shift patterns, and chain shifts are often being reversed. A common initiating feature appears to be TRAP lowering, which initiates a pull chain effect on the short front vowels, though LOT appears to also be an important factor.

1.5 SUMMARY AND THESIS OUTLINE

This chapter has introduced my study, and has provided the theoretical underpinnings of studies in chain shifts and mergers. It has furthermore provided a brief literature review on chain shifts and mergers present in various varieties of English, particularly ETEs. Via the description of the southern British English and the ETEs, this chapter provides an overview of global trends in vowel shifting to which SAfE can now be compared. Chapter 2 covers SAfE and provides information about its history, the various socio- and ethnolects as well as a general phonetic overview. Chapter 3 outlines the methods employed in this study and documents the use of Forced Alignment and Vowel Extraction (FAVE) and statistical methods, since this is the first study to use these particular packages in analysing SAfE. The results are presented in Chapters 4 and 5, which is followed by discussion and conclusions in Chapter 6.

CHAPTER 2: SOUTH AFRICAN ENGLISH

In the previous chapter South African English was characterised as an extraterritorial, Southern Hemisphere English. This chapter provides a more detailed account of the variety's development and the characteristics of its vowels. I begin with a discussion of the various models used to describe the formation of English in a colonial setting (section 2.1), after which the socio-historical background and the growth of local South African Englishes is described in detail (section 2.2). Regarding the phonetic features of SAfE (section 2.3), the classic work of scholars such as Hopwood (1928), Lanham (1967), Lass (1990, 2002, 2004), and others is outlined. Studies on SAfE have recently taken an acoustic turn, and those that draw on acoustic methods are discussed in section 2.3.3. The body of work on SAfE, which forms the basis of this study, will become familiar via these discussions and descriptions.

2.1 THE DEVELOPMENT OF (POST)COLONIAL ENGLISHES

In studying changes to varieties of English on a global level, the way in which the dialects in question arose could provide important insights. More specifically, knowledge of the ways in which SAfE as a colonial English developed and what the influencing factors were could shed light on the current similarities (and differences) between Englishes of a similar origin, and the Englishes of Britain as the source of the linguistic input. Regarding the evolution of (post)colonial Englishes in general, various approaches have been devised, chief among which are the Dynamic Model (Schneider 2003, 2007) and New-Dialect Formation (Trudgill 2004).

The approach that has received the most attention and critique from particularly South African scholars is Schneider's Dynamic Model. Generally, it is favoured over the approach by Trudgill (2004) in that Trudgill dismisses the role of social factors, particularly identity, in new dialect formation, rendering the process mechanical and inevitable (Van Rooy 2010, 5-6). On the other hand, the model put forward by Trudgill (2004) is particularly useful in discussing not only the historical formation of SAfE regarding input, which Schneider's model lacks, but also potential reasons behind various global changes observed of late. Though I discuss both models below, the work of Trudgill (2004) is most useful in explaining and documenting the current and previous characteristics of the phonology of SAfE, and so better suits the purposes of this study.

Schneider's Dynamic Model, put forth in his 2003 and 2007 work, presents a way of describing the evolution of Postcolonial Englishes. The underlying premise of the model, built on various others, is that all emerging varieties of English or New Englishes (in postcolonial settings) develop and evolve along the same underlying principles and processes, caused by the relationship between the two parties involved in colonial settlement (Schneider 2011, 34). In representing these underlying evolutionary processes, the Dynamic Model includes five developmental stages. Along with each of the stages, Schneider (2007) provides a description of various constituent conditions, which include 'history and politics', 'identity construction', 'sociolinguistics of contact/use/attitudes', and lastly, 'linguistic developments/structural effects'. I outline the model below, drawing from Schneider (2011, 34-5). I provide brief descriptions of the five stages in addition to the linguistic and structural effects of each at the expense of the other constituent conditions.

1. Foundation: English is brought to a new country via colonial enterprise, resulting in the emergence of bilingualism. During this stage, koinéisation takes place along with the borrowing of toponyms. In colonies used for trade, emerging pidginisation is also a feature. Koinéisation is the process whereby "a fairly neutral compromise dialect devoid of strongly dialectal forms" comes to exist (Schneider 2011, 240).
2. Exonormative stabilisation: The colonising country is politically and linguistically dominant when the colonial situation is stable. Linguistic norms are therefore representative of and in a way prescribed by the colonising country. 'Elite' bilingualism spreads among some members of the indigenous population. During this stage there is lexical borrowing of particularly cultural terms and names for vegetation and wildlife. In addition, pidginisation (or creolisation in trade and plantation colonies) occurs.
3. Nativisation: Bilingual speakers create a new variety of English that is heavily based on transfer, both phonological and lexical. This is as a result of the decrease in contact between the settlers and their home country, in conjunction with increased contact with other and local ethnicities. Conservative speakers dislike this new variety. Linguistically, this stage is characterised by 'heavy' lexical borrowing, phonological innovation (including accent stabilisation), structural nativisation, lexical productivity and code-mixing.

4. Endonormative stabilisation: once the colony gains independence from the colonial power (possibly after ‘Event X’), a need for nation building inspires the increased recognition of a new linguistic norm.¹⁶ This linguistic norm becomes codified and socially accepted. In this phase there is stabilisation of the variety, homogeneity is important, and dictionaries and grammatical descriptions are produced.
5. Differentiation is a stage that may occur. “In a stable young nation, internal social group identities become more important and get reflected in the growth of dialectal differences” (Schneider 2011, 35). Linguistically, this is the stage of dialect birth, where L1 or L2 ethnic, regional and social dialects develop.

Though many scholars have tested and applied the model to various (post)colonial Englishes, Schneider (2011, 35) notes, “the description abstracts from realities which in individual cases are more complex (for example, subsequent phases typically overlap, and not all constituent phenomena can be observed).”¹⁷ Regarding the application to English in South Africa, a number of scholars have critiqued the model, though it is generally accepted as very useful. Though Van Rooy (2010, 16) prefers Schneider’s model to Trudgill’s (2004) approach, he notes that it could be improved with more in-depth consideration of the input. Similarly, Coetzee-Van Rooy (2014, 54) suggests that the model would be even more useful after a “refinement of the identity construction for contexts similar to that of South Africa.” Bekker (2009, 86) suggests that the model would fit the South African situation better if (a) there was a less rigid adherence to the temporal sequencing, and (b) adaptation of the approach to different varieties, allowing each to develop at their own speed.

Schneider (2007, 188) contends that English in South Africa has progressed well into Stage 4, after democratisation in 1994 as the relevant Event X. Numerous scholars have disputed this, particularly given how far into Stage 4 Schneider suggests SAfE is (Bekker 2009, 86). Van Rooy and Terblance (2010), via a corpus consisting of English words collected from the newspaper *Vaal Weekly*, found evidence that SAfE is only starting to progress into Stage 4 (though without morphological derivations). Spencer (2011) investigates the extent to which Black South African English (BSAfE) has progressed along the model’s phases. She finds preliminary evidence (via acceptability ratings of entrenched features of BSAfE) that the

¹⁶ ‘Event X’ is defined as an event that causes mass disassociation, politically and in terms of individual identity, with the colonising country, which is replaced by association with the local (Schneider 2007, 48).

¹⁷ For responses to and applications of the model, see in particular Buschfeld, Hoffmann, Huber and Kautzsch (2014).

variety is only progressing *towards* Stage 4 (Spencer 2011, 147). Van Rooy (2014, 34-5) supports these findings, stating that though endonormative stabilisation has not been reached in South Africa, increasing acceptance of varying linguistic norms indicates a move into Stage 4. He argues however that endonormativity, though not unlikely, will take time to situate itself firmly into the South African community. In the same vein, Coetzee-Van Rooy (2014, 54) concludes that South Africa will very likely not progress past Stage 3 categorically.

Trudgill (2004), alternatively, argues for a deterministic, mechanical, predictable process of new-dialect formation for Southern Hemisphere Englishes, specifically considering the linguistic input. He bases his theory on a sample of speakers drawn from the extensive Origins of New Zealand English (ONZE) Project, and extends it to the other Southern Hemisphere Englishes (mostly SAfE and Australian English) given the similar nature and time frame of their colonisation by Britain. The ONZE database consists of 325 speakers, of which 84, born between 1850 and 1899, were used for his analyses. He believes that this time frame is the crucial period for the development of New Zealand English (Trudgill 2004, x-xi).

For Trudgill (2004, 23), the most important aspect of new-dialect formation is dialect contact. Based on the thorough analysis of his ONZE data and evidence found in the other colonial Southern Hemisphere Englishes, Trudgill (2004, 13) suggests that these varieties are “new and distinctive varieties of the English language which arose as a result of dialect contact, dialect mixture, and new-dialect formation.” For countries in the Southern Hemisphere, colonisation introduced varieties of English from the British Isles. He furthermore suggests that his thesis is supported by the similarities between New Zealand and Australian English, and South African and Falkland Islands English. Importantly, “they arose from similar mixtures in similar proportions occurring at similar times” (Trudgill 2004, 20).

An important aspect of Trudgill’s argument is the *tabula rasa* context in which new-dialects develop. Trudgill (2004, 26) defines *tabula rasa* situations as “those in which there is no prior-existing population speaking the language in question, either in the location in question or nearby.” Given this unusual dialect contact situation in colonial settings, Trudgill (2004, 27) argues that the new-dialect formation process is deterministic:

We can confidently predict that colonial dialect mixture situations involving adults speaking many different dialects of the same language will eventually and inevitably lead to the production of a new, unitary dialect.

Trudgill (2004, 27) posits two reasons for his deterministic approach to new-dialect formation: (1) the need to ‘talk like others talk’ (Keller 1994), similar to Jakobson’s (1971) phatic function, which is the desire to conform to social norms; and (2) crucially, accommodation. Usually, the process of accommodation is not difficult for children under the age of 8 (the critical period): in a new environment, they speak like the others speak. In a colonial situation where the speech of the others is far from uniform, the situation is more complicated. Trudgill (2004, 28) argues, based on the ONZE data, that accommodation in these settings will take 50 or so years (two generations), and that the most complicated process leading to a new, single norm is undertaken by children under the age of 8. He further notes that, in colonial, *tabula rasa* situations, social factors such as prestige do not influence children, which results in the similarity of outcomes from similar mixtures and hence illustrate a deterministic, mechanical process.

New-dialect formation consists of six processes, which are grouped into three stages indicative of the chronological, roughly generational time frames involved. My description of these processes and stages draws on those provided in Trudgill (2004, 84-115). The six processes are as follows:

1. Mixing. This involves the coming together of speakers of different dialects of the same language in the same geographical location.
2. Levelling. Here, variants that are in the demographic minority are lost. Importantly, Trudgill (2004, 85) notes that this is not an unsystematic process, and that it is not reliant on social factors. Simply put, the features that are in the demographic majority will remain at the expense of those in the demographic minority.
3. Unmarking. This is a subtype of levelling, where some variants survive despite being in the demographic minority. This happens when these minority variants are either unmarked, or are more regular than a majority variant (i.e. when two out of the three input dialects prefer a certain feature).
4. Interdialect Development. This occurs when certain variants arise out of the contact between the various dialects, but which are not features of either of the input dialects. This is a kind of compromise, where variants are linguistically intermediate between

two (or more) of the majority variants of fairly similar demographic strength. These intermediate forms are often simpler and more regular than the forms found in the input dialects. This process can also be seen as hyperadaptation via hypercorrection.

5. Reallocation. In cases where, after levelling, one of the competing variants survives in addition to another variant. Where these variants were once regional features, they become reallocated as social, stylistic or allophonic features.
6. Focussing. This is the process by which the new variety acquires norms and stability. Though it implies levelling, the reverse is not true.

Koinéisation takes place during the first five processes, which together with process 6 forms new-dialect formation.

Stage 1 of the new-dialect formation process involves rudimentary levelling and interdialect development. The rudimentary levelling process refers to the first contact and dialect mixing experienced by adult immigrants into the colonies, speaking various regional dialects of British English. This initial contact occurred at the assembly points in the British Isles, on the long boat journey to the colony and during the early days in the colony. Though many of the immigrants into New Zealand were agricultural workers who spoke Traditional-dialects, there is little evidence of 1800s Traditional-dialect features in the ONZE data, as well as in modern Englishes in New Zealand, Australia and South Africa. Trudgill (2004, 91) argues that, though Traditional-dialects were surely present on the boats, their features were levelled out in the early stages of dialect contact due to their minority status, largely influenced by normative attitudes and negative perceptions of rural dialects. An example of this rudimentary levelling is the lack of /v~/w/ merger in Englishes in New Zealand, South Africa or Australia. The merger involves /v/ and /w/ realised as /w/, so *village* is pronounced *willage* (Trudgill 2004, 91).

Interdialect development occurs at the early stage of dialect contact, and can result from accommodation and, potentially, hyperadaptation. A frequent feature of hyperadaptive interdialect development in the ONZE dataset is hypercorrective *h*-insertion in the initial position in words such as *and*, *I*, and *apple* (Trudgill 2004, 95). This process only occurs in Stage 1, and only in the speech of adults. Many of these interdialectal features will not survive beyond Stage 1.

Stage 2 of new-dialect formation is crucially focused on the role of children, and involves extreme variability and apparent levelling. The children born in the colonies (the second generation) are exposed to a chaotic mixture of features as a consequence of the dialect mixing of the first generation. As a result of being surrounded by various dialect features in the speech of their parents and other adults, these children have the opportunity to choose which features to adopt. This freedom of choice results in a high level of linguistic variability during this stage, which has three forms. The first form occurs when children choose features from the diverse set to which they are exposed, and new combinations are created. These combinations are high in number and individualistic, rather than regular in the speech of the society as a whole. In the ONZE database, Trudgill (2004, 103) finds that individual combinations are prevalent in the speech of those born between 1850 and 1870. These combinations, however, do not permanently affect modern New Zealand English (and by extension, the other Southern Hemisphere Englishes).

The second form of variability in the second stage of new-dialect formation involves intra-individual variability. Idiolects formed in dialect-mixture situations may be much more variable than idiolects formed in stable speech communities. The third form of variability involves inter-individual variability, where two speakers may differ from each other remarkably as far as dialect is concerned, even if they grew up together. This is as a result of the high levels of variability to which they are exposed, and their freedom to choose variables. In this sense, Trudgill (2004, 106) argues that Stage 2 is characterised by variable acquisition, not accommodation.

Another characteristic of Stage 2 is apparent levelling. Though Stage 2 is characterised by immense variation, Trudgill (2004, 109) argues that this variation is less than what was the case during early Stage 1. Therefore, some kind of levelling in addition to rudimentary levelling must have taken place. He supports this claim via the ONZE data, in which there is no evidence of certain widespread British English features. One such missing feature is a short vowel system consisting of only five vowels, where the FOOT vowel is used for both FOOT and STRUT (Trudgill 2004, 109). Since these features were surely present on the boats, their non-existence in the ONZE corpus suggests that they were lost. Trudgill (2004, 110) contends that these features did not survive into Stage 2 because of apparent levelling, not accommodation. Levelling is as a result of accommodation, where 'apparent' levelling suggests that the relevant feature was never acquired. In this sense there is a difference too

between rudimentary and apparent levelling. Essentially, the features that were demographically dominant were acquired, and those in the minority were not.

Stage 3 is the arrival at a final, stable, relatively uniform dialect, and is characterised by determinism. This stabilisation process occurred in the speech of New Zealanders who were born around 1890 (Trudgill 2004, 113). The stable, crystallised variety is as a result of the focusing process, as set forth by La Page and Tabouret-Keller (1985). Essentially, the koinéisation process involves the survival of variants in the majority. Stage 3 is characterised by far less dialect variation than Stages 1 and 2, and children therefore select the most common variants. Trudgill (2004, 115) argues that this selection process is not random but deterministic. Therefore, the phonological typology of New Zealand English (and by extension SAfE and Australian English) characterises it as a ‘southeast-of-England type’ because individual forms found in the southeast of England were also very often the majority forms in the original dialect mixture, not merely because many settlers came from there. Important to remember, however, is that southeast English dialect forms that were in the minority, such as *h*-dropping, did not survive, hence the hyperadaptive *h*-insertion during Stage 1 (Trudgill 2004, 115).

Trudgill’s (2004, 125-6) theory of new-dialect formation, based on New Zealand English and extended to SAfE and Australian English, can be summarised as follows: “new-dialect formation is deterministic in the sense that the outcome is predictable from the input.” The most common variants at Stage 1 were those most often selected at Stage 2 (despite variability stemming from freedom of choice). These most common variants were therefore the ones to survive into Stage 3, the new dialect called New Zealand English. Accommodation is crucial in Stage 3; during the koinéisation process, the majority wins.

Trudgill (2004, 129) argues that the formative period of the Southern Hemisphere Englishes is about 50 years. He therefore suggests that New Zealand English was formed between about 1840 and 1890. SAfE is a slightly earlier variety, formed by those born between 1820 and 1870 (corroborated by Lanham 1996 and Branford 1994), such that the first group who spoke a focused variety (amidst regional variation) consisted of adolescents in about 1885 (Trudgill 2004, 24). However, in the midst of the developmental period, there is regular language change, which potentially manifests in two ways. The first is the case where (post)colonial Englishes undergo language change experienced by the input dialects at an earlier stage (i.e.

delayed language change); and the second is the parallel development of global Englishes (i.e. simultaneous language change). These scenarios are referred to as ‘colonial lag’ and ‘drift’ respectively.

In colonial, *tabula rasa* situations, children are influenced by the dialects of their parents and other adults more than would usually be the case. As a result of this, colonial Englishes often experience colonial lag, which Trudgill (2004, 24) defines as follows:

(...) a lag or delay, which lasts for about one generation, in the normal progression and development of linguistic change, and which arises solely as an automatic consequence of the fact that there is often no common peer-group dialect for children to acquire in first-generation colonial situations involving dialect mixture.

Changes evident in Southern Hemisphere Englishes occurring after the same changes in the input varieties exemplify colonial lag. Trudgill (2004, 67ff) uses the case of rhoticity in British and Southern Hemisphere Englishes to exemplify this notion. Rhoticity was a common feature in Britain during the 1800s, which is reflected in the speech of 92% of the speakers in his ONZE database, who are all variably rhotic. That New Zealand English (and by extension, SAfE and Australian English) is today largely non-rhotic despite the rhoticity of the early speakers exemplifies colonial lag. Trudgill (2004) argues that New Zealand English inherited a ‘process whereby rhoticity is lost’ rather than non-rhoticity. Hence, the loss of rhoticity occurred in Britain a generation or so earlier than it did in Southern Hemisphere Englishes.

When there are parallel changes to Englishes around the world, these are instances of drift (Trudgill 2004, 130). In explaining drift, Trudgill (2004, 131) notes that the term was developed by Sapir (1921), and refers to inherent or inherited tendencies in languages and language families. He quotes Sapir (1921, 172), who notes, “drift is often such that languages long disconnected will pass through the same or strikingly similar phases.” Trudgill (2004, 131-2) suggests that drift is also exemplified in modern English: “language varieties may resemble one another because, having derived from the same common source, they continue to evolve linguistically in similar directions by undergoing similar linguistic changes.”

In Southern Hemisphere Englishes, Trudgill (2004, 142) identifies two kinds of drift. The first occurs when a linguistic change has already begun in Britain, which continues even after separation to the colonies. A change in progress during the 1800s in Britain was the Diphthong Shift, which the Southern Hemisphere Englishes inherited and continued (expanded on below). The continuation of the Diphthong Shift involves the increase in distance between the onset and offset of upgliding diphthongs, such as the shift from [aɪ] to [ɑɪ] in words such as *price* (Trudgill 2004, 139). This continuation occurred in parallel, though not necessarily at the same pace or to the same extent in each location (Trudgill 2004, 133).

The second type of drift occurs when there is no linguistic change in progress in the source dialect, though the varieties which stem from these common sources inherit shared tendencies and propensities, which potentially lead to the development of new changes even after separation (Trudgill 2004, 132). This kind of drift is more readily exemplified than the first type, with examples present in and shared by every Southern Hemisphere variety, barring SAfE in one instance only. This indicates that the Southern Hemisphere Englishes inherited propensities rather than ongoing changes (Trudgill 2004, 136). Essentially, certain parallel changes on-going in various Englishes are inexplicable via British antecedents (Trudgill 2004, 143).

An example of this kind of drift is happy tensing, which is a characteristic feature of Southern Hemisphere varieties. This involves the realisations of the final syllable in words like *happy* and *money* with /i/ rather than /ɪ/. Though this was rare in mid-1800s Britain, it is currently rapidly becoming the norm. Happy tensing is therefore a parallel development between the Southern Hemisphere Englishes and British Englishes over the last 150 years (Trudgill 2004, 136). Another example of this kind of drift is glide weakening, where Trudgill (2004, 139) notes that Southern Hemisphere Englishes inherited the propensity to decrease the distance between the onset and the offset of PRICE via the Diphthong Shift. British English accents lag in adopting this process, showing that propensity for change indicates only the possibility, not certainty that a change will occur. Southern Hemisphere Englishes therefore inherited the drift.

These two concepts, together with Trudgill's (2004) general model of dialect formation, provides a solid basis from which to discuss the development of SAfE as a (post)colonial English, in addition to providing the tools to aid in the analysis and description of changes apparent in SAfE and other global Englishes.

2.2 THE SOCIO-HISTORICAL DEVELOPMENT OF SOUTH AFRICAN ENGLISH

Bearing in mind Trudgill's (2004) model presented in section 2.1, this section provides a detailed, chronological description of the development of English in South Africa specifically, bringing various social factors, which are excluded by Trudgill (2004), to light. Spencer (1971, 8) writes that English was first heard on the African continent during the 1530s, when William Hawkins the Elder passed through *en route* to Brazil. Trade had begun during the 1500s and British ships sailed along the Guinea coast in West Africa (Schmied 1991, 6). Before the 1600s, the Cape served as a layover for European ships sailing to the East Indies (Mesthrie 2012a, 2093). During trade with the Khoesan inhabitants at the Cape, a lexicon consisting of Dutch, Portuguese and English developed (Den Besten 1989). Some Khoesan locals were taken to England to learn English, and others were taken to the East by ship. It was intended for those travelling East to serve as interpreters after learning English by way of force (Malherbe 1990). *Couree home go* ('Couree wants to go home') is the earliest (1613) recorded English utterance by a South African, showing the influence of Khoe SOV syntax (Den Besten 1989).

Since these English speakers did not settle in the Cape, this early contact with English was informal and irregular (Mesthrie 2012a, 2093). Only in the late 1700s did English begin to get a foothold in South Africa, and Lanham (1996) describes the process as occurring in three stages: (1) British settlement in the Cape between 1795 and 1820; (2) British Settlement in Natal in 1843, and (3) the influx of settlers due to the discovery of diamonds and gold during the 1870s and the 1880s. Figure 2.1 is a map of South Africa illustrating the stages of settlement as well as the political components of South Africa in the 1800s.

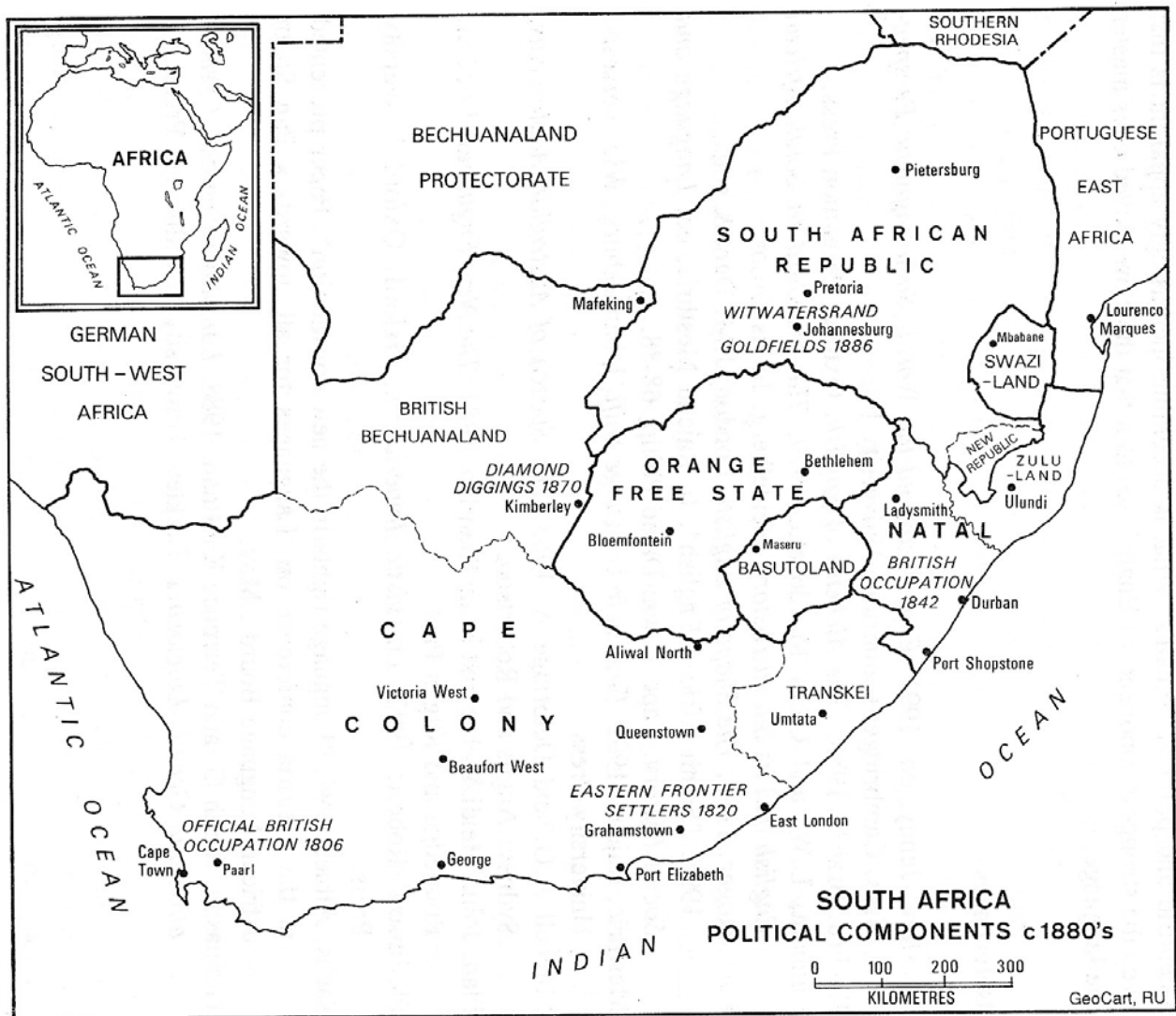


Figure 2.1: Map of South Africa's political components in the 1800s.
 Source: Lanham (1996, 18).

In 1795 Britain overthrew the Dutch-controlled Cape, though they did not initially establish a settlement. In 1802 the Dutch regained control, but again lost it to the British who reinvaded in 1806. At the 1814 Congress of Vienna, the European authorities handed the Cape over to Britain officially, and their governmental presence remained largely in the Western Cape (Lass 2004, 37). In Cape Town during this time, English was spoken by those involved in administration, sailing or the military (Mesthrie 2012a, 2097).

In 1818 the British parliament planned a settlement in the Eastern Cape (Lass 2004, 37). More than 5000 settlers arrived from various parts of Britain in 1820, constituting the first permanent English speaking settlement in South Africa. The settlers in the Eastern Cape were

largely from London and the Home Counties (including Buckingham, Hampshire, Sussex etc.), and consisted of people of mostly lower middle and working class backgrounds (Mesthrie 1993, 27).¹⁸ Lanham and Macdonald (1979, 73), like Trudgill (2004), note that distinct features of the English of these areas (and thus in settler speech) fell out of use early on. These include *h*-dropping as well as ‘*g*-dropping’ in words ending with *-ing*. The first locally born generations quickly lost contact with Britain, and were faced with poverty and continuous battle with the Xhosa community on the frontier, which lasted until 1878 (Lanham 1996, 20). Descendants of the settlers thus formed a frontier society as envisaged by the British parliament in 1818 (Lass 2004, 37).

Lanham (1996, 20-21) notes that the settlers were largely remoulded as a society due to poverty and the frontier conflict. This involved social levelling, leading to homogenous speech forms in the community within one or two generations. The children of the 1820 settlers thus spoke the first L1 variety of South African English. Contact with Dutch and isiXhosa resulted in a highly local dialect difficult for outsiders to understand (Lanham 1996, 20-21). Despite these local influences, Eastern Cape English was most similar to the two other Southern Hemisphere English varieties, found in Australia and New Zealand at the time, given that the 1820 settlement was only 32 years after the first British settlements in Australia, and 20 years before the settlement in New Zealand (Mesthrie 2012a, 2097).

In 1822, shortly after the first settlers arrived at the Cape, the governor, Lord Charles Somerset, declared English as the only official language of the Cape Colony, thereby exerting power over the Dutch (Mesthrie 1993, 27). Attempts to anglicise the long established Dutch community led to the first L2 variety of English in South Africa: Dutch English (Lanham 1996, 20).

In 1843 the British annexed the Boer republic Natalia (or Natal, today KwaZulu Natal) and established a second English-speaking settlement. Over 4000 immigrants arrived in Natal between 1848 and 1851/1862 (Mesthrie 1993, 27; Lanham 1996, 21). The Natal settlers were different to the Eastern Cape settlers in that they were mostly from higher social classes, and hailed from different areas of Britain, not just London and its surrounding counties. In

¹⁸ Bekker (2012a, 129) notes that this social class classification possibly needs revision, given this quote from Welsh (1998, 127) regarding the selection criteria for settlement in South Africa: “[they] were rather too strict, in that whilst they produced a high proportion of educated and responsible citizens, there were too few of the labourers and artisans needed for the pioneering work.”

addition, they were neither as impoverished, nor did they face frontier wars. According to Lanham (1996, 21), where the English in the Eastern Cape homogenised, the English in Natal did not: the social distinctions based on rank and position remained. Natal English was therefore the second L1 variety of South African English to emerge.

The Natal settlement was made up of a larger portion of settlers from the Midlands, Yorkshire and Lancashire, contributing to the differences between it and Eastern Cape English (Mesthrie 1993, 28). Moreover, there was contact with isiZulu rather than with isiXhosa, and very little with Dutch (Lanham 1996, 21). Where the English in the Eastern Cape was similar to Australian and New Zealand English, Natal English was perhaps more similar to English spoken in the southern United States, where settlement began in the 1600s. These two varieties were ‘closer’ to one another in that they shared prominent phonological features, such as glide weakening of /aɪ/ and fronting of /u:/ (Mesthrie 2012a, 2097).

During the 1860s and the 1870s, diamonds were discovered in Kimberley, and gold at the Witwatersrand, which encompasses Johannesburg and surrounding areas (Mesthrie 1993, 28). Between 1875 and 1904, 400,000 fortune seekers descended on the minefields from the Eastern and Western Cape, Natal and Britain (Lanham 1996, 22). In addition, Lanham (1982, 327) also mentions immigrants from eastern and western Europe. These European migrants were largely Jewish speakers of Yiddish, which Bekker (2009, 72) notes constituted the first “un-ancestral population to be integrated into the [SAfE] speech community.”¹⁹ The discovery of precious minerals led to large-scale changes to the socioeconomic climate in South Africa. Small, rural towns largely gave way to mining-industrial cities, with wealth and social status emerging as important aspirations. Cape English and Natal English were now in contact with other British English varieties, both social and regional (Lanham 1996, 22). In the mines and mining towns English became dominant with little opposition from Dutch. Standard Southern British English became the prestige variety, associated with high social status. This resulted in stigmatisation of the ‘local’ in favour of the quality and excellence of all things ‘British’. Towards the end of the 1940s, English was spoken in most cities (excluding Bloemfontein and Pretoria), larger towns and Natal, and less widely in the Western and Eastern Cape (Lanham 1996, 22).

¹⁹ In his discussion he draws on Kaplan and Robertson (1991).

There is some debate surrounding the impact of the mining era on local South African English. Lass (2004, 37) contends that the “discovery of minerals and the subsequent settler influx seems to have had little effect in general on the development of [SAfE].” Lanham and Macdonald (1979) argue that the Cape and Natal varieties were distinct and became more similar during the mining era. The reason they put forward is that the Eastern Cape settlers were lowest in social standing given their lower class composition upon immigration, promulgated in the mines by their low level of education. In contrast, the Natal settlers were generally of a higher social class, and had authority over the Eastern Cape settlers in the mines because they were more ‘English’ in that their ties with and experience of Britain were more recent (Lanham 1996, 23). The use of prestigious Natal variables thus increased in an attempt to use more prestigious, though less accessible, Southern British English. When white speakers of English returned to their homes from the mines, a more homogenous English spread (Lanham & Macdonald 1979).

Bekker (2009, 73ff), on the other hand, argues that Johannesburg was and is relatively independent from the other centres. As a *tabula rasa* context for dialect mixing, koinéisation and new-dialect formation, he posits Johannesburg as the context where SAfE emerged as a ‘late 19th century variety’, importantly different from colonial Cape and Natal English (Bekker 2012a, 133). Bekker (2009, 73) views the settlement of the Cape and Natal as providing “the main input into the development of a distinctly South African accent” and the fortune-seeking immigration to Johannesburg during the late 1800s “as the primary determinant in the development of social class dialects.” In light of this, Bekker (2009, 2012a, 2013) developed the Three-Stage Koinéisation Model of the formation of SAfE (figure 2.2).²⁰ This model represents the three stages in which SAfE became a koiné. It furthermore considers influences from other languages such as Yiddish and Afrikaans/Dutch.

²⁰ See Bekker (2009, 70-80; 2012) for the full rationale behind the model.

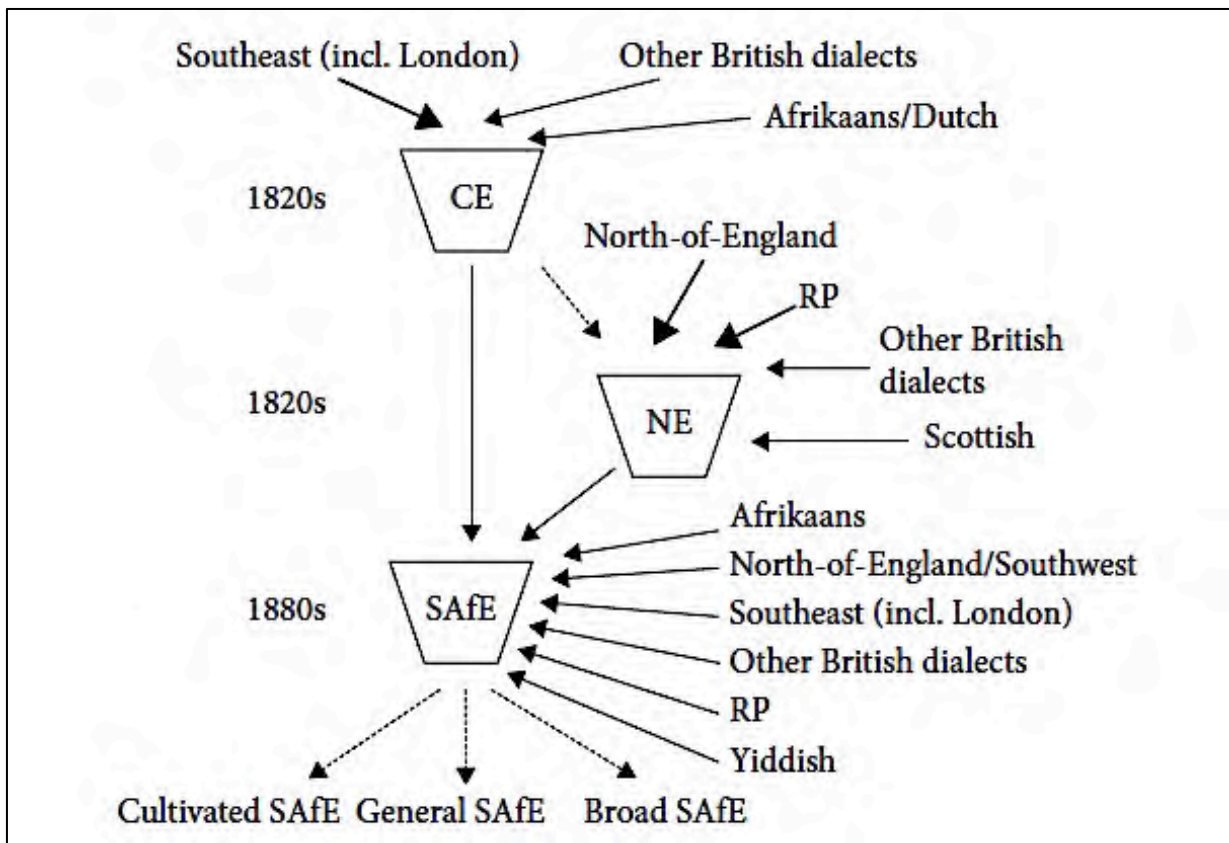


Figure 2.2: Bekker's Three-Stage Koinéisation Model of the development of SAfE (Bekker 2012a, 136).

In explaining the model presented in figure 2.2, Bekker (2012a, 135) notes that each trapezium represents a stage in the koinéisation process. Stage 1 is Cape English (CE), Stage 2 is Natal English (NE), and Stage 3 is Johannesburg (SAfE). The arrows are weighted differently, meant to visually articulate an estimation of the strength of each input variety's influence. In this sense, varieties of English from southeast England, as argued for by Trudgill (2004), had the strongest influence on Stage 1 of the process. Varieties from the north of England together with RP were the major influencers in Stage 2. The first two stages both influence Stage 3, though on the same level as other input varieties. These three stages together are thus seen as culminating in SAfE and its three sociolects: (1) Conservative SAfE, the least distinguishable from Southern British English, generally associated with older speakers who are in the middle to upper middle classes and are few in number; (2) General SAfE, which is seen as the local standard, associated with those in the middle class; and (3) Broad SAfE, very similar to Afrikaans English, generally associated with the lower middle

and working class.²¹ General and Broad SAfE are largely seen as ‘SAfE proper’, given the larger proportion of speakers (Lass 2002, 114).

Bekker (2012a) ultimately uses this model to show that SAfE is an extraterritorial variety of English that developed in the *late* 1800s, essentially springing from the contact situation in Johannesburg. This is contrary to other descriptions that posit SAfE as emerging in the early or mid 1800s, developed mostly from Cape Town English, Eastern Cape English and Natal English (such as Lanham & Macdonald 1979; Trudgill 2004). Bekker (2012a, 135) stresses that this is a working model open to critique and revision. It nevertheless provides a neat, concrete visualisation of the stages in which SAfE was influenced by other varieties of English and indeed other languages. What it perhaps lacks is information on the contact of English (especially in the Cape) with African languages like isiXhosa, particularly given the explicit mention made of this contact by Lanham (1996). Bekker (2012a, 132) however discounts any influence of African languages on SAfE beyond loan words, citing Schneider’s (2007, 175-81) model of the development of English.

Over and above the formation of L1 Englishes in South Africa, the 1800s saw the development of various L2 SAfE varieties in addition to Dutch English (Afrikaans English). Black South Africans had access to English in missionary schools in the Cape and Natal (Mesthrie 1993, 28). South Africans later classified as coloured largely spoke a local variety of Dutch, though English became a noteworthy second language around 1806 due to trade and employment (McCormick 1989, 73). Indian migrants arrived in Natal as indentured labourers in 1860, and learned English in addition to their mother tongues through schooling and via contact with English plantation owners. Within a few generations, Indian South Africans shifted to English as L1 at the expense of Indian languages, which survived mostly as heritage languages (see Mesthrie 1992).

Lanham (1996) identifies two further eras of importance to the subsequent development of SAfE in addition to the first three waves of settlement. The first is the ‘Post-war era’, spanning from the end of World War Two until about 1989 (Lanham 1996, 26). Fundamental to the development of English was the relationship between speakers of Afrikaans and English. In 1910, the local variety of Dutch (officially renamed Afrikaans in 1924) was given

²¹ See Lanham and Macdonald (1979) and Lass (2004) for in-depth descriptions. Early descriptions of these sociolects used ‘Cultivated’, ‘Respectable’ and ‘Extreme’ as labels. I use the less laden terms ‘Conservative’, ‘General’ and ‘Broad’ throughout this thesis.

official status alongside English via the Act of Union, which united the former Boer republics with the Cape Colony and Natal. Towards the mid 1940s, militant Afrikaner nationalists promoted hostility towards the socially and economically dominant English. This led to fierce language loyalty, resulting in what Lanham (1996, 25) described as the “deepest social division in white South African society.”

The Afrikaner Nationalist Party came into power in 1948, a few years after the end of World War Two, marking the start of Lanham’s ‘post war’ era. The Afrikaner Nationalist Party downplayed English in an attempt to redress the inferior status of Afrikaans. This led to a decrease in the frequency and proficiency with which Afrikaans people spoke English, and thus a decrease in Afrikaans-English bilingualism (Lanham 1996, 26). The National Party used education to elevate Afrikaans above English, and soon after legalising racially segregated schooling via the Bantu Education Act of 1953, implemented a mother-tongue instruction policy, replacing English as medium of instruction in black schools.

The white English-speaking community responded to National Party rule in various ways. English-Afrikaans bilingualism increased as a necessity, and there was re-distribution in the population as many English speakers relocated to the cities from towns (Lanham 1996, 28). Importantly, many rejected Afrikaner nationalism, and between 10 and 20 per cent of English speakers at the time rejected the Afrikaans nationalist view of South Africanism (Schlemmer 1976, cited in Lanham 1996, 28). In 1961 the Union of South Africa officially cut ties with Britain, becoming the independent Republic of South Africa, a move that arose out of Afrikaner nationalism rather than South African English speaker sentiments.

The effects of apartheid on the linguistic nature of South Africa are vast. In particular (as mentioned in Chapter 1, section 1.2), SAfE came to be characterised largely by ethnicity given its segregated development, promulgated by the Group Areas Act of 1950 as well as the Bantu Education Act.²² The rule of the National Party and segregation ended in the 1990s, starting a new era that Lanham (1996, 30) calls ‘The New South Africa’. The incoming democratic government, headed by the African National Congress, instituted a new constitution recognising eleven of South Africa’s languages as official; English and Afrikaans included (Lanham 1996, 30). Present-day SAfE (encompassing all varieties) has

²² The Group Areas Act specified racially segregated areas of residence and schooling, *inter alia*.

been subject to many changes due to the abolishment of racially based social engineering and segregation.

Similar to Conservative, General and Broad SAfE, Lass (2004, 374) notes that other L1 Englishes in South Africa, namely Indian South African English and Coloured South African English, have internal structures and differentiations particular to each variety. He further notes that speakers of these varieties may, in post-segregation South Africa, cross over into the linguistic space generally associated with White South African English (WSAfE), quoting Mesthrie (1995) and McCormick (1995) to this effect. This crossover has been supported in recent studies of SAfE, particularly Da Silva (2007), Mesthrie (2010), Wilmot (2014), Mesthrie, Chevalier and Dunne (2015) and Mesthrie, Chevalier and McLachlan (2015), and is discussed in section 2.3.

Given the various historical developments in the different parts of England during the time spanning from about 1300 to 1800 (which necessarily includes the period of British colonial expansion), Lass (2004) notes that Extraterritorial Englishes share certain features. The historical developments gave rise to various linguistic features, which allows the characterisation of certain English varieties as ‘southern’ (Lass 2004, 368). The term ‘southern’ does not necessarily refer to the Southern Hemisphere, but instead to the notion that these varieties are ‘of southern British type’. English spoken in the United States of America, South Africa, Australia and New Zealand fall into this category. Set out by Lass (2004, 368-9), these features are as follows (bearing in mind that post-millennium changes are not reflected here):

1. The diphthongisation of the Middle English /u:/ to /au/ for the lexical set MOUTH. In certain American varieties (particularly constrained to Pittsburgh²³) and in South Africa, this monophthongised to /a:/.
2. The lexical set GOAT became a reflex of Middle English /ɔ:/ rather than Middle English /a:/. The realisations of GOAT thus ranged from [o:] or [ou] to frontier [œu]. In some cases, secondary monophthongisation occurred, i.e. [œ:]
3. Due to the raising of older /a/ during the 1600s, the TRAP lexical set was realised as [æ] or higher. In Southern Hemisphere Englishes (i.e. SAfE, Australian English and New Zealand English) it raised further toward [ɛ].

²³ With thanks to Christopher Stelluf

4. The lexical sets STRUT and FOOT became phonemically distinct. FOOT became [ʊ], and STRUT varied between [ʌ], [ɐ], [ä] and [ĕ].
5. In most southern Englishes (and therefore absent in Northern England) the BATH and TRAP vowels became distinguished from other another (which is today a common feature). A typically southern English would therefore have realised TRAP with a short vowel and BATH with a long vowel, both with different qualities. Australian and New Zealand English realised TRAP with [æ] or [ɛ], and BATH with [a:]. South Africans mirrored these trends with the exception of [ɑ:] for BATH. In certain American varieties, length became the distinguishing feature between TRAP and BATH: the former lax and the latter tense.²⁴
6. A further feature common in most southern Englishes is to lengthen /æ/ before voiced stops and nasals (excluding /ŋ/). Therefore, /æ/ is the norm in words such as *bat*, *back* and *mat*, and /æ:/ in words like *bad*, *bag* and *man*.²⁵

Within this ‘southern’ type of English, there is a further division between the Southern and Northern Hemispheres. The Northern Hemisphere Englishes are largely ‘American’, whereas the Southern Hemisphere Englishes are more ‘British’ (as discussed in Chapter 1, section 1.4). Southern Hemisphere Englishes are, as a result of their histories, prone to the development of three major lectal types, which are socially hierarchically ranked (or perceived as such). This is particularly true of Englishes in South Africa (i.e. Conservative, General and Broad SAfE), New Zealand and Australia (Lass 2004, 372; Mitchell & Delbridge 1965; Lanham & Macdonald 1979).

²⁴ Lass (2004) terms this phenomenon ‘Lengthening I’, and dates it to the 1600s.

²⁵ Lass (2004) terms this phenomenon ‘Lengthening II’.

2.3 DESCRIPTIONS OF SOUTH AFRICAN ENGLISH

This section provides a brief overview of various phonetic descriptions of and acoustic studies on the vowels of SAfE. It focuses on the seminal work of early scholars and also provides information on important, more recent acoustic studies. The discussion begins with a brief outline of the work of various scholars who have contributed to the field, followed by the description of the monophthongal system, particularly as described in the classic and non-acoustic sources. Thereafter the acoustic studies are outlined and compared to these earlier works.

English as has been described for white and black South Africans forms the basis of the discussion below, though much work has also been done on the other varieties. Coloured South African English has been described by Wood (1987), Finn (2004), Mesthrie (2010, 2012a, 2012b), Brown (2012) and Toefy (2014). The seminal description of Indian South African English has been the work of Mesthrie (1992, 2002, 2012b, 2014), and other works include those by Delbridge (2006); Wiebesiek (2007); Chevalier (2011) and Mesthrie and Chevalier (2014).

Phonetic descriptions of the WSAfE vowel system are numerous. Hopwood (1928) provides the earliest compilation of ‘South African English Pronunciation’. While very detailed, he often does not overtly distinguish between L1 and L2 varieties (Bekker 2009, 142). Nevertheless, Hopwood (1928) provides valuable insights into the English spoken in South Africa in the early 1900s, on which many modern transcriptions rely as a point of comparison.

Lanham was a prolific researcher into the phonetic system of WSAfE between 1960 and 1980, producing a significant body of work, often in collaboration with other scholars (such as the edited volume Lanham & Prinsloo 1978). Lanham and Traill (1962) and Lanham (1967) improved on Hopwood’s (1928) descriptions by focussing only on L1 speakers of English in South Africa, particularly speakers of the ‘prestige’ or ‘educated’ dialect. They therefore describe the Conservative and General SAfE of the time (Bekker 2009, 145-9). Importantly, Lanham (1978) and Lanham and Macdonald (1979) were responsible for the first descriptions correlating social factors to phonetic features, namely the lectal continuum of Conservative, General and Broad SAfE.

Another prolific SAfE scholar was Lass (1990, 1995, 2002, 2004), whose auditory descriptions of WSAfE are highly influential and comprehensive. His works are often cited as a starting point to studies on the variety. Of particular value to this thesis is the collaboration between Lass and Wright (1985, 1986), describing the Older Short Front Vowel Shift.

In addition to these scholars whose focus fell on South Africa in particular, scholars such as Wells (1982) are often cited in relation to SAfE in that they provide broad descriptions of numerous Englishes around the world. In addition to Wells (1982), scholars such as Hooper (1945, 1951, 1952), Breckwoldt (1961), Branford (1994) and Bowerman (2004) are good descriptive sources of the phonetics of WSAfE.

According to Van Rooy (2004, 943), “research on [Black South African English] has understandably had a predominantly pedagogical bias.” Research into the phonetic system of the variety was only recently brought into focus, though Hundleby (1964) is an early source (Van Rooy 2004, 944). Regarding the phonetic system, much work has been done on the influence of the L1 on L2 BSAfE, including the work of Hundleby (1964), Adendorff and Savini-Beck (1993), and Van den Heever and Wissing (2000). Van Rooy and Van Huyssteen (2000, 30) note of this research and their own that there is little evidence of differentiating influence on English by the different mother tongues of the speakers.

Where WSAfE is socially characterised by what Lass (2002) calls the Lectal Trichotomy, BSAfE is distinguished in terms of a lectal continuum imported from Creole Studies by Mesthrie (1997). BSAfE can therefore be described as basilectal, acrolectal or mesolectal. Van Rooy (2004, 947) focuses on the latter two, and explains that “the acrolect is closer to native varieties of SAfE” than the mesolect. Acrolectal BSAfE is characterised by variability when compared to the mesolect, though the most noteworthy feature is the use of both tense and lax monophthongs, largely absent in the mesolectal vowel system. Mesolectal BSAfE makes use of five vowels only, all lax: /i/, /ɛ/, /a/, /ɔ/ and /u/ (Van Rooy 2004, 946). Acrolectal speakers use the central vowel space, producing vowels like /ə/ and /ʌ/. In addition, the reduction of unstressed vowels in acrolectal BSAfE occurs in ways very similar to WSAfE (Van Rooy 2004, 947).

Research primarily focussing on BSAfE pronunciation have been provided by Van Rooy (2000); Van Rooy and Van Huyssteen (2000) and Van Der Pas, Wissing and Zonneveld (2000). These works are largely subsumed in Van Rooy (2004), a comprehensive descriptive characterisation of BSAfE, drawing on these descriptions in addition to an acoustic vowel analysis of mesolectal BSAfE provided in Van Rooy and Van Huyssteen (2000).

Da Silva (2007, 239) in her doctoral study of white and black South African students at the University of the Witwatersrand in Johannesburg concludes that there are two lects represented by her citation-form data. White South Africans mainly speak 'Lect 1', and its features correspond to those of traditional WSAfE of the 'General' variety. 'Lect 2', largely spoken by black South Africans, contains new variants that are not found in the literature on traditional BSAfE (Da Silva 2007, 245). Speakers of 'Lect 1' are "more selective and absolute regarding the choice and distribution of variants" and most of the variants that appear in 'Lect 1' also appear in 'Lect 2' (Da Silva 2007, 160-161). These varieties, once quite separate, now share features. Of particular relevance are her findings for KIT, STRUT, and NURSE, which are included in the descriptions to follow.

2.3.1 CLASSIC DESCRIPTIONS OF THE MONOPHTHONGAL SYSTEM

The work of the oft cited and classic scholars, particularly Hopwood (1928), Lanham (1967), Webb (1982), Lass (1990, 2002, 2004) and Van Rooy (2004), as an overarching descriptive source, form the basis of the descriptions to follow. In particular, the discussion provides information on vowel variation within the three sociolects of WSAfE in addition to the mesolectal and acrolectal varieties of BSAfE. Each vowel is dealt with on an individual basis, after which vowel chain shift patterns are discussed.

KIT

Wells (1982, 612-3) identifies a singular feature of SAfE, which has come to be known as the KIT split, where the vowel is split between fronter and more central realisations. When KIT occurs around velars, word initially and after /h/, the realisations are more front; the centralised realisations occur elsewhere. In General SAfE the KIT split is present as [ɪ] and [i̠]. In Broad SAfE there is a greater distance between the two allophones, with the fronter allophone at [i] and the centralised allophone at [i̠]. Speakers of Conservative SAfE, on the other hand, are not regular users of the KIT split, and use [ɪ] as the sole realisation of KIT (Lass 2002, 115; Lass 2004, 375). In General and Broad SAfE, KIT retracts before /l/, and can occur as far back as [ɨ̠], leading to merger or near-merger with FOOT (Lass 1990, 275; Lass 2002, 115). According to Van Rooy (2004, 947), the KIT split does not occur in BSAfE. Acrolectal speakers favour [ɪ], though [i] is also heard. Mesolectal speakers use only [i]. Da Silva (2007, 157), on the other hand, identifies a centralised [ə] in addition to [ɪ] for speakers of Lect 2. Though she does not overtly mention the presence of a split, this is perhaps indicative of the emergence of a KIT split in the English of black South Africans.

DRESS

According to Lanham (1967, 81), the DRESS vowel is realised as [E], which in his descriptive system is a mid vowel occurring midway between [e] and [ɛ]. The General and Broad realisations of DRESS in more modern WSAfE is fairly high [e], and Lass (2002, 115) notes that women tend to have higher realisations of DRESS than men. Some varieties of General SAfE and Broad SAfE raise DRESS to [e̠] or [ɪ] (Bowerman 2004, 937). Lass (1990, 276) furthermore notes that DRESS is often centralised, for women in particular, and is high enough therefore to approach /i/. Before final /l/, DRESS undergoes extreme lowering and retraction to

/ɛ/ or /æ/ (Lass 2002, 115). Social and gender variability is not a feature of DRESS in BSafE: [ɛ] is favoured by both acrolectal and mesolectal varieties (Van Rooy 2004, 945-7).

TRAP

Given the process of Lengthening II, the TRAP vowel is perceptibly lengthened before voiced codas, so that *mat* and *mad* are realised as /mæt/ and /mæ:d/ respectively. Lass (2004, 376) describes this vowel as “highly variable”, best characterised as between cardinal [a] and cardinal [ɛ], though closer to [ɛ], in the speech of Conservative and General speakers. Importantly, Lass (2002, 115) states that the vowel is never realised as [ɛ] by these speakers, though it approaches it. The [ɛ] realisation is a feature of Broad SAfE, and speakers may further raise TRAP above [ɛ], which is stereotyped as characteristic of working-class speech (Lass 2002, 115; Lass 2004, 376). A fairly raised TRAP vowel is also indicated by Hopwood (1928, 12) who provides [ɛ(˘)] as the main variant, with a more retracted [ɞ] occurring before nasals.²⁶ Interestingly, this value for TRAP is exactly what he posits for DRESS, indicating overlap between the vowel categories (Hopwood 1928, 10). In acrolectal BSafE there is free variation between [æ] and higher [ɛ], though mesolectal speakers prefer [ɛ] (Van Rooy 2004, 945-7).

STRUT

Lanham (1967, 82) describes this vowel as “unrounded, central [ʌ]” or slightly lower. Lass (1990, 277) on the other hand provides a wide range of STRUT realisations, from [ä] to [ä̃], and suggests that a reasonable symbol to denote this range is /ɐ/. Later, Lass (2002, 115) notes that the more retracted variants of STRUT are associated with older General and Conservative speakers, and describes the norm for General SAfE as a range between [a̠] and [ä]. Over time, therefore, STRUT in WSAfE has become progressively more fronted, though in certain idiolects, STRUT may retract before /l/ towards [ɒ] (Lanham 1967, 82). In BSafE STRUT is a back vowel [ʌ], though at times a more central and lower [ü] is heard. The use of [ʌ] as a phoneme is an acrolectal feature of BSafE and is not found in mesolectal BSafE, where [ä] is preferred (Van Rooy 2004, 945-7). Da Silva (2007, 157) identifies low front [a] as the realisation most common for her black speakers, suggesting a much fronter vowel than [ʌ] or [ü].

²⁶ Hopwood’s (1928, 11) symbol is [a], which he describes as being ‘low and middle’. Bekker (2009, 143) notes that [ɞ] is a more appropriate symbol.

LOT

Hopwood (1928, 8) describes the most frequent realisation of LOT as back, short rounded [ɔ]. Similarly, Lass (2002, 115) notes that LOT is typically a back vowel (though often centralised), and weakly rounded. He therefore sets the norm for LOT at [ɔ̄]. In General SAfE, he notes a tendency towards unrounded, central [ä] among young speakers in Cape Town and KwaZulu Natal (Lass 2004, 376). Generally speaking, therefore, LOT appears to be a fronting vowel in WSAfE. In BSAfE there is free variation between fully back [ɔ] and [ɒ] for acrolectal speakers, while the mesolectal speakers use only [ɔ] (Van Rooy 2004, 945-7).

FOOT

The earliest description of this vowel is [ú(·)] (Hopwood 1928, 10), which can be represented by [ʊ] in modern IPA (Bekker 2009, 143). Lanham (1967, 82) describes this vowel as ranging between fairly low /ɜ/ and /o/ (represented in modern IPA by /ɵ/ and /ʊ/), and it therefore has a centralising tendency. Lass (1990, 277) confirms this centralisation, and provides [ö] as the norm for FOOT. This holds for all speakers of WSAfE, though Lass (2002, 117) notes a trend among young General SAfE speaking women to front it to somewhere just below [ʊ]. In acrolectal BSAfE [ʊ] is preferred, though [u] is also used. Mesolectal speakers only use [u] (Van Rooy 2004, 945-7).

FLEECE

This vowel is largely stable in WSAfE, with long, high front [i:] being the realisation for all speakers (Lass 2004, 376). Hopwood (1928, 12) similarly notes [í] or [i:]. A common feature in the other Southern Hemisphere varieties is the diphthongisation of FLEECE (and GOOSE), which is absent in SAfE (Lass 2004, 376). In BSAfE there is variation between [i] and [ɪ], though [i] is more frequent in acrolectal speech. Mesolectal speakers do not show this variation and prefer [i]. Furthermore, Van Rooy (2004, 947) notes that, in contrast to the mesolect, acrolectal BSAfE sometimes differentiates KIT as lax and FLEECE as tense.

NURSE

Hopwood (1928, 11) described WSAfE NURSE as varying between rounded, low mid [œ(r)] and mid high, more central [ë:], though the latter is infrequent. Lass (1990, 278) describes this vowel as one that is always rounded, except in the most Conservative varieties, and transcribes the usual realisation as [ø:], though some speakers lower it to [œ:]. Lass (2004, 377) notes that unrounded realisations of NURSE emerge in the speech of young women who

received their education at private schools (discussed in section 2.3.3). In BSAfE there is a range of realisations between unrounded [ɜ] and [ə], and sometimes [ɛ]. This variation is an acrolectal feature of BSAfE not found in mesolectal BSAfE. Mesolectal speakers use only [ɛ] (Van Rooy 2004, 945-7). Da Silva (2007, 157) identifies central [ə] and front [ɛ] as the realisations for NURSE most common for her black speakers. This perhaps indicates that [ɛ] has become a more stable feature in this variety of SAfE.

GOOSE

As with many other varieties of English (such as southeast British English) this vowel has undergone a process of fronting in SAfE. Speakers of Conservative SAfE, however, retain an RP-like [u:] realisation (Bowerman 2004, 937). For General SAfE, Lass (1990, 278) describes GOOSE as ranging from [u:] to [ʊ:], and even [y:] in particularly female speech, though he later extends it to all young people: “the younger the speaker, the fronter the realisation” (Lass 2004, 377). He furthermore notes that fronted GOOSE is perceived as “particularly ‘white’”, in contrast to speakers of other varieties, such as Coloured and Indian SAfE, who have much backer realisations. In Van Rooy’s (2004, 947) description, GOOSE in BSAfE is realised as [u] or slightly lower [ʊ], mostly short.

BATH

[including PALM and START (Wells 1982, 615)]

The earliest description by Hopwood (1928, 8) identifies low back rounded [ɒ] as the most definite realisation of BATH, with a more central [ɐ(r)] as the chief variant. According to Lass (1990, 278), “the typical range is from advanced [ɑ:] to centralised [ä:].” Slightly later, Lass (2002, 116-117) connects BATH with social variability: in Conservative SAfE, BATH realisations range from [ä:] to [ɑ:] in ‘posher’ styles. In General SAfE this vowel is backer, and can reach [ɑ:] in the speech of particularly men and younger speakers. Lanham (1967, 91; 1978, 153) notes that, in Broad SAfE, BATH is rounded and retracted, so that *pot* (from the LOT set) and *part* (from the BATH set) are minimal pairs showing only a distinction in vowel quantity (*pot* is short, *part* is long). Lass (1990, 278) confirms this, and suggests that the range of BATH realisations for Broad SAfE is (almost cardinal) [ɑ:], /ɒ:/ and sometimes [ɔ:]. Van Rooy (2004, 947) describes BATH as ranging between [ä] and [ʌ] in both mesolectal and acrolectal BSAfE.

THOUGHT
[including NORTH and FORCE (Wells 1982, 615)]

Hopwood (1928, 8) notes variation between mid to low [ɔː(r)] and raised [oː(r)]. This variation (and raising) has stabilised, and generally, THOUGHT is realised as mid high [oː] in General and Broad SAfE (Lass 1990, 278). Conservative SAfE speakers may realise this vowel in the vicinity of [ɔː], though this is no longer common (Lass 2004, 337). In BSAfE the vowel is [ɔ] which is sometimes tense, but mostly lax in both varieties (Van Rooy 2004, 947).

2.3.2 THE OLDER SHORT FRONT VOWEL SHIFT

When compared to other varieties of British English, Lass (2002, 113) notes that SAfE has higher DRESS and TRAP realisations, along with a “very centralised” KIT nucleus. Various reasons have been postulated for this. Lanham and Macdonald (1979, 46) are unsure as to the origins of the raised DRESS and TRAP variants, though they attribute centralised KIT to contact with Afrikaans. Lass and Wright (1986) argued contrastively that the ‘atomistic’ borrowing of a single vowel feature is unlikely. They instead suggest that the raised short vowels are connected in the form of a chain shift. Called the South African Chain Shift, it is “a nineteenth-century vowel shift in which raising of /æ/ toward [ɛ] and raising of original /ɛ/ to [e] seems to have forced (most of) original /ɪ/ to centralise” (Lass 2002, 113). Table 2.1 provides a summary (Lass & Wright 1986) of the comparison between SAfE and RP by Lass and Wright (1985, 137ff), illustrating the comparative raising (and/or centralisation) of the short vowels in SAfE.

Vowel	SAfE	RP
KIT	ĩ	ɪ
DRESS	e ~ ě	ɛ
TRAP	ɛ	æ

Table 2.1: Comparison between KIT, DRESS and TRAP in SAfE and RP.
(Lass and Wright 1986, 208).

The raising of the short front vowels, together with the centralisation of KIT, appears to be the end result of a process involving a very complex and variable system (Lass 2002, 113-4). The development of the South African Chain Shift is presented in detail in Lass and Wright (1986) and is summarised in Lass (2002, 114ff). These scholars draw, in particular, on the writings of an 1820 settler, Jeremiah Goldswain, for phonetic clues pertaining to the English of that time. In his largely non-standard spelling there are examples of, to quote Lass (2002, 114):

Raised TRAP: *contractor*, ‘contractor’; *atrected*, ‘attracted’; *lamentation*, ‘lamentation’

Lowered DRESS: *amadick*, ‘emetic’; *hadge*, ‘hedge’; *sant*, ‘sent’

Raised DRESS: *git*, ‘get’; *kittle*, ‘kettle’; *liter*, ‘letter’

Lowered KIT: *presner*, ‘prisoner’; *deferent*, ‘different’; *sleped*, ‘slipped’

Retracted KIT: *buld*, ‘build’; *busket*, ‘biscuit’; *contunerd*, ‘continued’

Lass and Wright (1986, 211) conclude that the stimulus of the South African Chain Shift, characteristic at that time of S AfE, can be sourced from the Midlands and the south of England. Though this kind of complexity and variation is not unusual in language development, particularly in rural dialects of mainland English, they eventually led to a chain shift in S AfE. “What seems to have happened is that over time the four categories involved ‘spaced themselves out’ by raising, and KIT became more and more centralised” (Lass 2002, 114). Lanham (1965) notes the centralisation of KIT, and predicts that it will continue.

Trudgill (2004), however, is of a different opinion. He notes that this kind of short front vowel raising has also been postulated for New Zealand English (Bauer 1979, 1992), and as is the case in S AfE, it is described as a push chain initiated by TRAP raising (Trudgill 2004, 42). The ONZE data contained raised DRESS and TRAP, much like what is shown in Jeremiah Goldswain’s writing. Trudgill (2004, 43) therefore argues that raised variants of these vowels were brought to New Zealand (and by extension South Africa) from Britain. British Englishes have subsequently innovated the lowering of the short front vowels, comprising the lowering of [e > ɛ] and [ɛ > æ] in a drag chain initiated by TRAP (Trudgill 2004, 42-43). Since New Zealand English has not experienced this lowering, Trudgill (2004, 43) suggests that the raised DRESS and TRAP realisations are conservative, indicative of the settler input prior to subsequent lowering in Britain.

Essentially, Trudgill (2004, 43) argues that Southern Hemisphere Englishes (excluding Falklands Island English) have close short front vowels “because they inherited these qualities from south-eastern English English, which subsequently distanced itself from them by acquiring more open qualities.” He further substantiates this claim by noting the greater probability of linguistic innovation occurring in one place, rather than the lesser probability of the same linguistic innovation occurring in four different, separated places at about the same time (Trudgill 2004, 43). The subsequent lowering of the short front vowels in

Australian English (dating to the 1960s), and by extension SAfE, can by his argument possibly be attributed to colonial lag.

2.3.3 THE ACOUSTIC ERA

The current trend among linguists, especially sociophoneticians working on English, is the acoustic analysis of vowels. An early acoustic study of SAfE was done by Webb (1983). Though his work provides interesting insights into the changing face of SAfE, it is slightly problematic, since the data consists of only 249 tokens gathered from his own speech (Bekker 2009, 150). Recent acoustic studies of South African Englishes are numerous, though the most prolific researcher into the variety of English spoken by white South Africans is Bekker, who has produced a significant amount of research since 2007. Regarding (acrolectal) English as spoken by black South Africans, important acoustic studies include the work of, among others, Morreira (2012) and Wilmot (2014). Mesthrie (2010), in addition to Mesthrie, Chevalier and Dunne (2015) and Mesthrie, Chevalier and McLachlan (2015), provides comparative acoustic insights into the speech of black and white South Africans (*inter alia*). These and other studies are outlined below, drawing attention to changes or updates in comparison to the earlier descriptions outlined in the previous section.

The acoustic work of Bekker and Eley (2007) is of particular relevance. Their dataset consists of ten young, white, middle class women, five from Johannesburg and five from East London. Each subject read two reading passages and a word list, and the analysis is based on 11 selected words from the word list. Their tentative conclusion suggests the emergence of a new prestige variable, particularly in the elite northern suburbs of Johannesburg: lowered and retracted TRAP, which is largely indistinguishable from STRUT (Bekker & Eley 2007, 113). This is a particularly important finding, as it provides the first evidence of lowering TRAP in SAfE.

Bekker (2009) produced a doctoral dissertation on the vowels of SAfE, which can be described as the first major acoustically based study of a variety of SAfE. In addition to a valuable extended critical review of various sources describing SAfE, he compares his data to Webb's (1983) acoustic study, despite its shortcomings. Bekker's findings are based on word list data elicited from young white women from various areas of South Africa. His findings largely corroborate the descriptions provided by the classic scholars, though they provide important additional information as follows:

1. The TRAP vowel is found to lower and retract, which, importantly, corroborates the findings of Bekker and Eley (2007). In light of this, Bekker (2009) suggests that (General) SAfE participates to some extent in the Southeast English Chain Shift (which is discussed in Chapter 1, section 1.4.1).
2. Bekker (2009, 357) infers from his data that STRUT in General SAfE is considerably fronted, as suggested by earlier scholars reviewed. He however contends that it is a potential change in progress, given comparison to Webb's (1983) data, and that the changes to TRAP and STRUT could culminate in either vowel merger or a repositioning of each lexical set.
3. Bekker (2009, 357-8) labels LOT an indicator of General SAfE given its stability as a relatively fronted vowel. This suggests that the fronting of LOT, as noticed by the classic scholars, has stabilised, and that the vowel is no longer fully back (especially in the speech of young women).
4. The fronting of FOOT is present in Bekker's (2009, 358-9) data, though he adds, importantly, that significantly fronted and lowered FOOT overlaps with KIT allophones before final /l/ (though FOOT itself retracts in this environment). Importantly, Lass (1990, 2002) notes this in particular reference to KIT before /l/, which he states overlaps with FOOT to a degree suggestive of merger or near-merger.
5. Though Lass (2004) notes some unrounding of NURSE in the speech of privately educated young women, Bekker (2009, 393) notes that the vowel does not vary to a large degree, and that it is always rounded.
6. The raising of THOUGHT as noticed by the classic scholars is confirmed in Bekker (2009, 312), whose results provide "some support for the notion that this vowel has undergone endogenous Pattern-3-like raising in General SAfE." In Pattern 3, back vowels front and other back vowels raise to take their place, as discussed in Chapter 1, section 1.3.2.

A particular finding of Bekker's (2009) work, which he expands upon in Bekker (2014a) is in relation to the KIT split. There has been some debate over the exact nature of the KIT split, and Bekker (2014a) questions what the term has come to represent in recent thinking, based on a thorough review of the descriptions by earlier scholars. Bekker (2012a, 143) notes that his 2009 work supports the notion that the split is not phonemic, as suggested by Wells (1982). Instead, he argues that it is established allophonic variation. Bekker (2014a, 129), in reviewing the literature surrounding the KIT split, and supplemented by acoustic analysis

concludes, “there is both impressionistic and acoustic evidence to suggest that this polarisation of qualities across phonetic space is not always in evidence in modern SAfE, and certainly not for modern [General SAfE].” He concedes, however, that the acoustic evidence he cites is based on citation-form data, necessitating investigations into other styles. In relation to Coloured South African English, Toefy (forthcoming) found a binary split between what she calls the IT and the SIT lexical sets. The IT set is realised frontier and higher towards [ɪ], where the SIT set is slightly lower and retracted towards [ə], indicating some polarisation.

Du Plessis and Bekker (2014) provide an interesting, though not acoustic, discussion of neorhoticity in SAfE. Their data consists of structured interviews with eight young, white women who read various passages before answering an adapted danger of death question to elicit more unmonitored speech. Du Plessis and Bekker (2014, 29) found high levels of rhoticity in the careful speech section of the interview, though not in the casual style section. In addition, the NURSE lexical set favours rhoticity the most (Du Plessis & Bekker 2014, 33). In comparison with work done by particularly Hartmann and Zerbian (2009), they suggest, despite tenuous data, that black middle class South Africans lead the “current trend towards neorhoticity” to which white speakers accommodate (Du Plessis & Bekker 2014, 32). Further acoustic studies on consonant variation in SAfE are Bekker (2007) on *s*-fronting in General SAfE; and O’Grady and Bekker (2011) on dentalisation as an indicator of regionality.

The masters dissertation of Wileman (2011) provides a regional investigation into General SAfE, with particular reference to PRICE, NURSE and KIT in the English of Durban and Cape Town. He found significant differences particularly in the trajectory of the glide of PRICE, concluding that glide-weakening is a feature of English in Durban more so than of English in Cape Town, particularly in female speech. While no regional differentiation was found for NURSE, KIT, in most environments, proves more centralised in Durban than in Cape Town English, particularly in the speech of men (Wileman 2011, 118).

Regarding the acoustic analysis of vowels across varieties of SAfE, Mesthrie has been a prolific researcher. In particular, he has noted the patterns of ‘crossing over’, where young, particularly black middle class speakers use the linguistic system formerly associated by both

the linguists cited above and non-linguists, with prestige (white) English.²⁷ In this vein, he documents the ‘deracialisation’ of the GOOSE vowel, where young white and black speakers were the most similar in their fronted realisations of the vowel (Mesthrie 2010). Mesthrie, Chevalier and McLachlan (2015) further support these findings via a perception test using the speech of black and white speakers of English. The findings show that, in the speech of young women in particular, it is no longer possible to easily gauge a person’s ethnicity based on their English accent, at least for the middle classes.

Mesthrie, Chevalier and Dunne (2015), through a thorough study into the variation of the BATH vowel across all four ethnicities in five South African cities, provide a detailed social and regional appraisal of the steady backing and raising of BATH in SAfE. Where Lanham (1978, 153) notes that the backing of BATH is a feature of Broad SAfE, Mesthrie et al. (2015, 26) find that this social restriction no longer holds. The ‘general’ speakers in their sample show a wide range of variants from [ɑ:] to [ɒ:] to [ɔ:], with men showing greater raising and backing than women.

The work of Morreira (2012) and Wilmot (2014) compare the speech of black and white South Africans to track post-apartheid social change and its influence on English. These two studies rely in particular on the schooling system as a variable. In South Africa, private schools are those where the fees are very high and no support is received from the government. As such, pupils attending private schools are often from wealthy households. As a result of segregated history, white children are in a majority at these schools, though many private schools did not exclude pupils from other ethnicities under apartheid law. Model C schools are those that receive funding from the government, and in times of segregation were reserved for white pupils. The fees for these schools are much less than for private schools, and the student body therefore consists of a broader spectrum of social classes than the private schools. Post-1994 these schools opened their doors to all South Africans, and have come to be termed former model C or ex-model C schools

²⁷ Mesthrie (2010, 13) notes that the term ‘crossover’ should not be confused with ‘crossing’ as used by Rampton (1995), which refers to the “playful and stylised use by young people of varieties not traditionally associated with them.”

Morreira's (2012) doctoral work analyses the accent and attitudes of 44 black university students with a model C schooling background. Her findings support the notion of crossover speech, though neither categorically nor consistently. Regarding the KIT vowel, for instance, Morreira (2012, 153) finds evidence of the KIT split, though only for some speakers, based on word list data. A feature more regularly shared with white speakers of SAfE is KIT retraction before /l/. Morreira (2012, 154) furthermore finds evidence of fronting FOOT, and suggests that this indicates potential accommodation to the centralised variant associated with white speakers of SAfE. Similarly, her findings indicate a fronted GOOSE vowel, particularly in the speech of young women, correlated with the length of contact with white speakers of SAfE (Morreira 2012, 186-7). Where Van Rooy (2004) describes LOT as mostly lax [ɔ], Morreira (2012, 155) finds realisations close to [o:], rather like the norm associated with white speakers.

Wilmot (2014) compares the speech of white and black schoolgirls in Grahamstown (Eastern Cape), particularly in respect of the type of school attended and mother tongue. She gathered data from a former model C school and a private school, and her sample consisted of L1 isiXhosa and L1 English speakers between the ages of 16 and 18. Morreira's (2012) data suggests that, though her speakers are similar to white speakers in certain ways, the extent to which schooling affects accent is not straightforward. Wilmot's (2014) findings on the other hand show that social class (or schooling) and mother tongue correlates strongly with aspects of crossover speech (or in her wording, accommodation), particularly with regards the DRESS, TRAP and GOOSE vowels.

Importantly, Wilmot (2014, 333) finds no significant differences between white and black private school girls for DRESS. Both groups produce a fairly low [ɛ]. She contends that the white speakers lower DRESS in response to international trends (i.e. short front vowel lowering) to which the black speakers accommodate. The former model C isiXhosa speakers have higher DRESS values, accommodating to the norms of raised DRESS as per the Older Short Front Vowel Shift (Wilmot 2014, 333-4). Mother tongue isiXhosa speakers who attend private schools are likely to lower TRAP to [æ], where the white L1 English speakers retract the vowel to [æ] or even [a]. The former model C isiXhosa speakers, on the other hand, retain more traditional variables, ranging between [ɛ] and [æ], whereas their white L1 English peers prefer a slightly retracted [æ] (Wilmot 2014, 332-3). GOOSE fronting is prevalent in the speech of the L1 isiXhosa schoolgirls, particularly in the coronal environment. Those in

private school environments produce realisations close to [u:] with little variation. isiXhosa speakers in a former model C environment show more variation, ranging between more traditional [o] and crossover, frontier [u:] (Wilmot 2014, 331).

The post-1994 sociolinguistic climate may be figuratively described as a gold mine for the sociophonetic enterprise. Indeed, changes in the South African context provide “near-laboratory conditions for the study of speech divergence and subsequent convergence” (Mesthrie 2010, 5). While much work has been done on SAfE, there seems to be a lack of comprehensive acoustic descriptions of the vowel space. Bekker (2009) is the only such description, though limited in the sense that it is based on citation form data produced by young white women. Given the labour associated with acoustic research, it is not surprising that many scholars choose to focus on a few vowel variables at a time. (This problem is largely remedied by the new methods discussed in Chapter 3.)

2.3.4 VOWEL MERGERS

Only two vowel mergers have been identified in SAfE: the Second FORCE Merger and Weak Vowel Merger. The Second FORCE merger involves the merger of CURE /ʊə/ with THOUGHT /o:/, where *cure* is pronounced [kjo:] (Bekker 2013b, 1). This merger is as a result of CURE lowering, the stages of which for RP were: [ʊə] > [oə] > [ɔə] > [ɔ:] (Wells 1982, 56-7). Bekker (2013b, 4) notes that the Second FORCE merger is present in Southern Hemisphere Englishes, though its presence is, importantly, partial and variable, as it is for RP. He also notes, citing Trudgill (2004), that the vowel is often pronounced [o:], and not [ɔ:] as in RP. While some words in the CURE class merge with THOUGHT, others merge with GOOSE and NURSE. In some cases therefore, *cruel* becomes /kru:l/ in SAfE (Lanham 1967) and *surely* becomes /ʃɜ:li/ (Bekker 2013b, 12).²⁸ The circumstances under which either GOOSE or NURSE occur is variable (see Bekker 2013b, 31-32). Trudgill (2004, 145) linked the Second FORCE Merger in the other Southern Hemisphere Varieties to drift, though he noted that it had not yet occurred in SAfE. Bekker (2013b, 33) concludes of his analysis that Second FORCE Merger, while variably present in his SAfE sample, is not spreading extensively. In this sense, though some SAfE speakers participate in the Second FORCE Merger, it is conditional at best, and has many exceptions (Bekker 2015, personal communication).

²⁸ Or narrowly, [ʃø:] for the first syllable (Bekker 2013b, 12).

Bekker's (2014b) acoustic analysis of the weak vowels (letter, comma, happy) and weak vowels occurring in places other than final position (in words like *abbot* and *Lennon*), finds evidence to support the Weak Vowel Merger in SAfE. Wells (1982, 167) describes Weak Vowel Merger as the loss of distinction between the final vowels in *Lenin* and *Lennon*, pronounced in RP as [lenɪn] and [lenən]. Bekker (2014b, 144) concludes that the Weak Vowel Merger in SAfE has taken place, though the merged quality is not /ə/ as it is in other varieties for unstressed vowels in final position, but rather something like [ə] or [ɨ]. Where RP has distinctions in the final vowels of *abbot* and *rabbit* ([ə] vs. [ɪ]), there is essentially no distinction in General SAfE (Bekker 2014b, 142).

2.3.5 THE REVERSE SHORT FRONT VOWEL SHIFT

Recently, the Older Short Front Vowel Shift appears to be reversing. Mesthrie (2012a, 2098) cites impressionistic evidence of particularly the lowering and retraction of TRAP, beginning around the year 2000 in the speech of young, middle class white women in Cape Town and Durban. Bekker and Eley (2007) and Bekker (2009), the first to actually study this reversal, confirm this lowering and retraction acoustically for female General SAfE speakers in Johannesburg. Bekker (2009) links it with the tendencies observed in the chain shift in operation in southeast England (Torgersen & Kerswill 2004). In addition to the lowering and retraction of TRAP, Mesthrie (2012a, 2098) also reports the emergent lowering of DRESS to [ɛ] and the lowering and centralisation of KIT to [ɛ̃], particularly in velar and glottal environments.²⁹

Mesthrie (2012a, 2099) offers various possibilities with regards to the causes or origins of the incipient reversal of the older shift. He first noticed these changes around the year 2000, a time in South Africa of socio-political change as well as the 'gap year' phenomenon, where many young white South Africans spent a year or more working in England before returning to begin their tertiary education. Linguistically, these young people were surrounded by qualities of KIT, DRESS and TRAP that were comparatively lower, which indicates a possible contact-induced change.

Mesthrie (2012a, 2099) cites two international, parallel developments that point to another possible origin for short front vowel lowering. Firstly, the lowering and retraction of TRAP in

²⁹ Though the centralization of KIT was predicted as a continuation of the Older Vowel Shift, it is seen as part of the Reverse Vowel Shift in general, as is discussed in Chapter 4, section 4.2.

Australia during the 1960s and 1990s (Horvath 2004, 640); and secondly, short front vowel lowering (and/or retraction) in Canada (Clark, Elms & Youssef, 1995). He argues that these parallel developments are in response to an ‘earlier prime mover’. He rejects RP as that potential prime mover as it is unlikely to have had a significant influence in Canada (see for example Chambers (2004)). Instead, Mesthrie (2012a, 2099) points to the short front vowel lowering and retraction in California, documented by Gordon (2004), *inter alia*.

Given the abundance of audio-visual entertainment from Hollywood (more popular than those of British origin) consumed by South African youth, it is possible that California (and thus ‘Valley-speak’) is a source for these changes to SAfE (Mesthrie 2012a, 2100), i.e. it could be a change from above. Although the likelihood of language change aided by the television is deemed by sociolinguists to be slight (Chambers 1998), the possibility should not be ruled out *prima facie* in an era of globalisation and mass satellite communication. Bekker (2009, 63; 437), though confirming this as a possibility, instead attributes it to the vernacular-driven shift in the south of England as a result of colonial lag. This thesis investigates the exact nature of the Reverse Short Front Vowel Shift in SAfE, 12 years after Mesthrie first noticed its emergence.

2.4 CONCLUSION

This chapter has provided a detailed account of the socio-historical development of SAfE as well as the phonetic qualities of the monophthongal vowel system. Importantly, it outlined the Older Short Front Vowel Shift, which is the starting point of the analyses to follow. The methodology employed in this study is outlined in the following chapter, after which the analysis is provided in Chapters 4 and 5.

CHAPTER 3: RESEARCH & ANALYTICAL METHODOLOGY

This chapter outlines the methodology employed in gathering and analysing the vowel data presented in this thesis. The considerations surrounding the data collection are discussed in section 3.1, after which section 3.2 describes the final sample in detail socio-demographically. The interview process is outlined in section 3.3. The Forced Alignment and Vowel Extraction Toolkit was used to obtain vowel data, and is discussed in section 3.4. Section 3.5 elaborates on the various statistical methods as well as the conventions adopted for the presentation of the results and analyses.

3.1 DATA COLLECTION

Judgement sampling was employed to identify potential participants. This method sees the researcher identifying, before the data-collection stage, the kind of speaker that would be ideal for inclusion in the study. Thereafter, the interviewer sets about completing a quota of interviews with speakers who fit the identified criteria (Milroy & Gordon 2003, 30). Sankoff (1980) outlines three important areas to be considered in specifying the criteria to be met by potential participants: the boundary of the speech community or group of interest, the stratification of the sample and the sample size (Milroy & Gordon 2003, 26). In addition to these considerations, the criteria guiding the data collection process are discussed below.

The speech community is restricted to Cape Town, located in the Western Cape Province of South Africa (figure 3.1). If a speaker was not born in Cape Town, they must have spent the majority of their lives (or at the very least, their entire high school career) in the area. Region is deemed important since Cape Town's suburbs (see figure 3.2) are not all English speaking. In order to sample from 'English speaking Cape Town', therefore, the speech community excludes predominantly Afrikaans speaking areas like Durbanville, Bellville, and Bloubergstrand, which are locally referred to as the 'Northern Suburbs' (indicated in figure 3.2). The remaining suburbs of Cape Town that are largely, though not exclusively, English dominant are locally referred to as (a) the 'Southern Suburbs', including Rondebosch, Newlands, Claremont, Kenilworth and Wynberg; and (b) the 'Deep South', including Muizenberg, Fish Hoek, Simon's Town, Noordhoek, Kommetjie and Scarborough. Areas in the city bowl and long the Atlantic seaboard, such as Sea Point, Green Point, Camps Bay, Milnerton and Table View, are largely English speaking.



Figure 3.1: Map of South Africa, showing the various provinces and prominent cities/towns.³⁰



Figure 3.2: Map of the suburbs and towns of Cape Town, South Africa.³¹

³⁰ Source: http://media.withtank.com/fc3336e41c/south_africa_province_map.png

³¹ Source: <http://www.roomsforafrica.com/dest/south-africa/western-cape/regions/cape-peninsula.jsp>

The sample is stratified by gender, age and ethnicity. The main focus of the analysis fell on the group most likely to use, or observed as users of the Reverse Short Front Vowel Shift. This group is largely young, white middle class speakers. In view of recent ‘crossover’ research, an additional aim was to see if the Reverse Short Front Vowel Shift is in use in other groups. To keep the project statistically tight, and in view of the country’s demographics, the analysis focuses on black middle class speakers. The age of the bulk of the participants was therefore limited to between 18 and 30 years; the age-bracket including speakers likely to show changes to the Older Short Front Vowel Shift, given Mesthrie’s (2012a) comments on the emergence of the Reverse Short Front Vowel Shift around the year 2000. A few older white speakers (i.e. over the age of 40) as likely participants in the Older Short Front Vowel Shift were also included for broad apparent time comparison (Gordon 2002, 258).

3.2 THE SAMPLE

The bulk of the participants were recruited via the internal mailing system of the Linguistics Section of the University of Cape Town (UCT). Additional participants were recruited via an approved campus-wide email (provided in Appendix A). Where necessary friends and colleagues were approached to either participate or to facilitate contact with potential speakers. All in all I conducted interviews with 72 volunteers between 2013 and 2015. Of these, only 51 interviews were used for analysis, for the following reasons:

1. The quality of certain recordings was not good enough for acoustic analysis (expanded upon in section 3.3). Four such recordings were excluded.
2. Some speakers who volunteered were either speakers of Afrikaans English, or had not received the majority of their schooling in Cape Town. Other volunteers were ethnically either coloured or Indian, and thus were not included in this particular study. Seventeen such participants were excluded.

In addition to the remaining 51 participants, two middle class black male speakers were added from Rajend Mesthrie’s database on South African English. Finding participants (especially English-dominant, middle class, black speakers) within the time frame proved difficult. The final number of speakers in the sample is thus 53, broken down into 44 white speakers (21 men, 23 women) and 9 black speakers (5 women, 4 men). The demographic information of the final sample is presented alphabetically in table 3.1.

n	Pseudonym	Birth Year	Gender	Ethnicity	Schooling	Tertiary Education
1	AaronA	1985	Male	White	Private	Master's degree
2	AlanL	1994	Male	White	FMC	Student
3	BobbyC	1983	Male	White	Private	Bachelors degree & diplomas
4	AnathiN	1993	Female	Black	FMC	Student
5	AndiswaM	1996	Female	Black	FMC	Student
6	BradP	1993	Male	White	FMC	Diploma, student
7	BrendaF	1956	Female	White	FMC	Matric
8	CaitlynS	1994	Female	White	FMC	Student
9	CeboM	1991	Male	Black	FMC	Master's degree
10	ChloëV	1994	Female	White	Private	Student
11	CindyH	1995	Female	White	FMC	Student
12	CraigS	1987	Male	White	FMC	Honours degree
13	EricaV	1993	Female	White	FMC	Student
14	GaryL	1991	Male	White	Private	Student
15	GeoffP	1990	Male	White	Private	Student
16	GlenD	1984	Male	White	FMC	Master's degree
17	HayleyS	1994	Female	White	Private	Student
18	IlseP	1993	Female	White	FMC	Student
19	JaneK	1989	Female	White	FMC	Diploma
20	JanetH	1992	Female	White	FMC	Student
21	JennyL	1994	Female	White	Private	Student
22	JerryM	1991	Male	White	Private	Bachelors degree
23	JimH	1981	Male	White	FMC	Student
24	KaleyW	1988	Female	White	FMC	Honours degree
25	KaraL	1990	Female	White	FMC	Honours degree
26	KayC	1961	Female	White	FMC	Student
27	LarryR	1993	Male	White	Private	Student
28	LauraE	1994	Female	White	Private	Student
29	LeeL	1994	Female	White	Private	Student
30	LexiK	1988	Female	White	FMC	Student
31	LiamH	1994	Male	White	FMC	Student
32	MandisaM	1994	Female	Black	FMC	Student
33	MarilynK	1987	Female	White	FMC	Bachelors degree
34	MikeF	1985	Male	White	FMC	Masters degree
35	MilesG	1991	Male	White	Private	Student
36	MindyG	1994	Female	White	FMC	Student
37	OlwethuN	1993	Female	Black	FMC	Student
38	PaddyF	1988	Male	White	FMC	Bachelors degree & diplomas
39	PartickM	1990	Male	White	FMC	Bachelors degree
40	PerryF	1954	Male	White	FMC	Bachelors degree
41	RachelJ	1994	Female	White	FMC	Student
42	RichS	1994	Male	White	FMC	Student
43	RuthW	1993	Female	White	FMC	Student
44	SamuelB	1988	Male	White	FMC	Bachelors degree
45	SandileQ	1993	Male	Black	FMC	Student
46	SarahN	1993	Female	White	FMC	Student

47	TamiH	1994	Female	White	FMC	Student
48	TedR	1994	Male	White	Private	Student
49	ThandiD	1994	Female	Black	FMC	Student
50	ThembaM	1992	Male	Black	FMC	Student
51	ThukileM	1988	Male	Black	FMC	Student
52	TodK	1991	Male	White	FMC	Student
53	ValV	1958	Female	White	FMC	Matric

Table 3.1: Demographic information of the selected sample of 53 speakers.

All participants in the selected sample attended either government/public (former model C/FMC) or private/independent (private) schools. This essentially places the participants in the middle class or the emergent middle class (who are in constant contact with the language used by the middle class via schooling). The majority of the participants are affiliated to UCT in that they are either undergraduate or postgraduate students, or recent UCT graduates in employment. Two speakers, BrendaF and ValV completed only high school (Matric) with no tertiary education. These speakers are both older and chose to stay at home with their children, which was common practice during the 1980s when their children were born.

Given the study's focus on potential users of the Reverse Short Front Vowel Shift, the majority of the speakers (98%, n=49) were born between 1981 and 1995 (i.e. between 18 and 32 years of age at the time of the interview). The remaining 2% (n=4) represents the older speakers, aged between 53 and 60 at the time of the interview (born between 1954 and 1961).

3.3 THE INTERVIEW

The majority of the interviews took place with students on the campus of UCT (n=46), and others were with people in employment in private homes (n=7). Each participant signed a consent form (Appendix A) and in each case, they were informed at the beginning of the interview that participation was voluntary, that they could withdraw at any point, and that their identity would be protected.

The methods associated with the well-known sociolinguistic interview (Labov 1972) were employed in a slightly adapted form. When recording participants with the intention of studying their language use, sociolinguistic researchers are faced with the observer's paradox: the problem of pursuing natural and spontaneous speech data in unnatural, unspontaneous settings. There are numerous ways in which the researcher can reduce the effect of observer's paradox, one of which is the use of an in-group interviewer:

(...) the more similar two speakers are, the less likely the interviewer effect/observer's paradox is to complicate the interview, the ideal situation is to have the interviewer/interlocutor (...) be from the same social group as the interviewee (Di Paolo & Yaeger-Dror 2011, 10).

Though some of the participants knew me as a tutor in the Linguistics Section, others had never met me before. Nevertheless, I had much in common with them: I am a student, a Capetonian, and I am within the age bracket of the younger group of participants. Despite regular social interactions with certain of the participants (my friends), the introduction of a recorder with the request of a natural conversation proved problematic. One participant was aware of the recorder throughout the interview, gave stilted responses and covered up the microphone with her hand each time she said something she did not want recorded.³² Conversely, another participant disregarded the recorder and microphone immediately and had a very emotional, personal conversation with me for over two hours. To avoid the possibility of future interference, subsequent interviews with friends were conducted with more than one speaker at a time, in line with what Di Paolo and Yaeger-Dror (2011, 11) suggest.

I took various steps to create an environment in which the participant felt comfortable, particularly when we were not in their homes. To ensure that the surroundings of my office did not induce formality, we sat on the couches rather than around the desk, often with a hot beverage beforehand to inspire a relaxed approach to the interview. The casual style portion of the interview was largely unstructured. The main aim was to elicit the vernacular, and therefore the topic and direction of speech was geared towards a conversation rather than a question-and-answer interview. In most cases, I would start with a question about their studies, after which the conversation developed naturally. In most cases the conversations centred on studies, work, interests and (social) life in Cape Town.

A common tool to elicit the vernacular is the danger of death question (Labov 1971), with which I did not have much success. Instead, I would nudge the participant into a narrative based on something that has come up in the conversation. This led to many extended, free-flowing discussions about favourite novels, movies, overseas trips and music, which the

³² In addition to her wariness of the recorder, the quality of the recording was unsuitable for acoustic analysis. This interview was therefore excluded from the study.

participants were excited about. In cases where I was unfamiliar with a novel or a movie, the participants would describe it to me. In many interviews the participants provided narratives without much direction. These narratives were of recent occurrences such as incidents concerning the building of a new home, a year spent in the army or overseas, or a disagreement with a close friend. In other cases, stories of inappropriate behaviour, bunking school or drug and alcohol abuse were told spontaneously. In many of the interviews the participants spoke with me for over an hour, which leads me to believe that they were comfortable and very likely to have used their vernacular.

The structured part of the interview included a word list and a minimal pairs test. The word list was used as a tool to elicit citation form style (though not used in the analyses to follow) and is based on Wells' lexical sets with some additions, notably a parallel list of words in a B_D carrier (Appendix B). The words were presented to the speaker as a slideshow on an iPad (with pauses between each slide) with the aim of avoiding list intonation (Bekker 2009, 124) as well as the rustle of paper. A minimal pairs test followed the word list, consisting of a list of minimal pairs of TRAP and STRUT, such as *putt, pat*; and *buck, back* (Appendix B). The aim of this test was to ascertain whether the production of these vowels matched their perception, a crucial step in studying vowel mergers (Gordon 2002, 247). The participants were asked to read each minimal pair aloud, and to judge whether the words sounded the same or different. This task was also presented on an iPad. After these tasks were complete, the participants were given the opportunity to ask about the particulars of the research.

The first few interviews were recorded on an Olympus DS5000 with a lapel microphone (Shure Lavalier Condenser, MX185), which unfortunately proved problematic. In particular, certain interviews were so soft that they were essentially inaudible. In another case, a cell phone was too close to the recorder, and though on silent, the signal waves interfered with the recording. The remainder of the recordings were made with a steady-state Marantz Professional (PDM661 MKII). This recorder is much better suited to interviews intended for sociophonetic analysis, since there is the option to control the recording settings to a much finer degree than on the Olympus. In addition, the quality of these recordings is infinitely superior. A few interviews were still recorded with the aid of the lapel microphone, though its use was very problematic and eventually it was abandoned in favour of the Marantz's internal microphone. The problems associated with the lapel microphone included accidental tugging on the chord, long, loose hair constantly rubbing on it, awkward placement of the microphone

based on what the participant wore and grainy recordings (some of which were excluded). In addition, and most importantly, the use of a lapel microphone undermined the pursuit of unmonitored speech, given the constant reminder of it on the participants' clothing. The majority of the interviews used in the sample were therefore recorded on the Marantz without a lapel microphone. The recordings were done in mono MP3 (320 bit rate, 44.1kHz sample rate), subsequently converted to 16bit WAV. Once the recordings were done, they were prepared for formant measurement.

3.4 AUTOMATIC FORMANT MEASUREMENT

Automatic vowel analysis is a fairly new analytical approach for South African varieties of English. Toefy (2014) pioneered its use in her study of Coloured South African English via the Penn Phonetics Lab Forced Aligner (P2FA). The Forced Alignment and Vowel Extraction suite (Rosenfelder, Fruehwald, Evanini & Yuan 2011) was written in the programming language Python, and is an adaptation of P2FA (developed by Yuan and Lieberman 2008). The suite consists of two toolkits. The first is FAVE-Align, written by Rosenfelder, geared specifically towards sociolinguistic interviews. The second is FAVE-Extract, an adaptation of *extractFormants* (Evanini 2009) by Rosenfelder and Fruehwald. FAVE was first used to analyse data from the Philadelphia Neighbourhood Corpus, and so was trained on American English data.³³ This section provides step-by-step documentation the use of the Forced Alignment and Vowel Extraction suite on South African English data, since this is the first study to do so.

3.4.1 INSTALLING FAVE

The Forced Alignment and Vowel Extraction suite (FAVE) is available via an online interface and is also downloadable for local installation from the website or from Fruehwald's Github.³⁴ For local installation (command-line versions), the suite requires a Macintosh computer, though steps have been taken more recently to make it compatible with Microsoft computers. Though the online interface is user-friendly and easy to access, local installation allows for greater control over analysis settings and the dictionary, both particularly important for first time use on SAfE.

³³ Information on FAVE can be found at <http://fave.ling.upenn.edu/about.html>.

³⁴ <http://fave.ling.upenn.edu/about.html> / <https://github.com/JoFrhwld/FAVE/wiki>

The installation of FAVE was a challenge, largely due to my inexperience with Macintosh computers and with computer programming. The zip-file download, however, comes with excellent step-by-step guides for the installation process, which eased the burden somewhat. Support for the installation and subsequent use of FAVE is provided in Fruehwald's Github, as well as in a Google Group for users of FAVE.³⁵ Some of FAVE's developers (particularly Joe Fruehwald and Keelan Evanini) and frequent users contribute to the discussions of problems, such that FAVE is continually being updated and debugged.

The FAVE process relies on an internal dictionary, comprised of words and their phonemic transcriptions. FAVE's default dictionary is the Carnegie Mellon University (CMU) dictionary, based on North American English varieties.³⁶ The phonemic transcriptions are therefore very different to what would be provided in a dictionary appropriate for SAfE. As a result, it was replaced with the British English Example Pronouncing (BEEP) dictionary (Robinson 1994), confirming the benefit of local installation.³⁷ The BEEP dictionary is preferable in that it is modelled on British pronunciations, which are more similar to SAfE than North American Englishes. In particular, the lack or near-lack of rhoticity in many non-American English varieties (such as SAfE) is accounted for.

The transcription system used by the CMU and BEEP dictionaries is called ARPAbet, an orthographic version of the IPA that FAVE has been trained to recognise. Table 3.1 shows the various ARPAbet symbols together with example words and corresponding transcriptions. Lexical stress is indicated on the vowel by a system of numbers. Number 1 indicates primary stress, 2 indicates secondary stress, and 0 indicates an unstressed vowel. The word *kit* is therefore transcribed, with spaces between each sound, as K IH1 T.

The ARPAbet symbol AH refers to both STRUT and schwa. Therefore, stress is used to tell these two vowels apart. STRUT is AH1, and schwa is AH0, though this requires manual editing in the output. In the CMU and BEEP dictionaries, LOT and BATH are both transcribed as AA. Since these are two different vowels in SAfE and other British Englishes, a different symbol, OH, is used to refer to LOT as separate from AA for BATH (following Baghai-Ravary, Grau & Kochanski 2011; Toefy 2014). This additional phoneme cannot be easily added to the

³⁵ <https://groups.google.com/forum/#!forum/fave-users>.

³⁶ <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

³⁷ <http://svr-www.eng.cam.ac.uk/comp.speech/Section1/Lexical/beep.html>

inner workings of the FAVE suite, however. The manner in which I deal with the complexities surrounding AA, AH and other vowels is expanded upon in section 3.4.4.

VOWELS			CONSONANTS		
Phoneme	Example	Transcription	Phoneme	Example	Transcription
AA	BATH	B AA TH	B	be	B IY
OH	LOT	L AA T			
AE	CAT	K AE T	CH	cheese	CH IY Z
AH	STRUT ABOUT	S T AH T AH B AW T	D	day	D EY
AO	THOUGHT	TH AO1 T	DH	they	DH EY
AW	MOUTH	M AW TH	F	fee	F IY
AY	PRICE	P R AY S	G	go	G OW
EH	DRESS	D R EH S	HH	he	HH IY
ER	NURSE	N ER S	JH	just	JH AH S T
EY	FACE	F EY S	K	key	K IY
IH	KIT	K IH T	L	late	L EY T
IY	FLEECE	F L IY S	M	me	M IY
OW	GOAT	G OW T	N	knee	N IY
OY	CHOICE	CH OY S	NG	sing	S IH NG
UH	FOOT	F UH T	P	pay	P EY
UW	GOOSE	G UW S	R	read	R IY D
			S	sea	S IY
			SH	she	SH IY
			T	tea	T IY
			TH	thanks	TH AE NG K S
			V	vein	V EY N
			W	we	W IY
			Y	yes	Y EH S
			Z	zoo	Z UW
			ZH	pleasure	P L EH ZH ER

Table 3.2: The ARPAbet symbols of transcription with corresponding lexical sets (source: <http://fave.ling.upenn.edu/usingFAAValign.html>).

Figure 3.3 shows the entries for the various versions of ‘ranch’ in the BEEP dictionary. The CMU dictionary transcribes this word as R AE1 N CH /ræ:ntʃ/, where the transcription in the BEEP dictionary is R AA1 N CH /rɑ:ntʃ/. Furthermore, the word ‘rancher’ is entered twice, once without rhotic /r/ and once with, for two reasons: (a) to take into account varieties of English that are rhotic (such as Irish English) and those that are not (RP or SAfE); and (b) to take into account the presence of linking-r in connected speech. The CMU dictionary has only one entry for this word, with an r-coloured vowel: R AE1 N CH ER1. Though the use of the BEEP dictionary is much preferred to editing the 134 000 words in the CMU dictionary to suit SAfE, it is not without problems. Certain differences between the vowels of British English and SAfE need to be changed for a South Africa-specific dictionary. I deal with these in section 3.4.4. Throughout the process of analysis, several additions and changes have been made to the dictionary. In total, 718 words were added.

RANCH	R	AA1	N	CH	
RANCH'S	R	AA1	N	CH	IH1 Z
RANCHED	R	AA1	N	CH	T
RANCHER	R	AA1	N	CH	AH0
RANCHER	R	AA1	N	CH	AH0 R
RANCHER'S	R	AA1	N	CH	AH0 R Z
RANCHER'S	R	AA1	N	CH	AH0 Z
RANCHERS	R	AA1	N	CH	AH0 Z
RANCHES	R	AA1	N	CH	IH1 Z

Figure 3.3: Example of words found in the British English Example Pronouncing dictionary (Robinson 1994).

The process of analysis using the FAVE suite is outlined in the remainder of the section. Descriptions of the inner workings of the suite can be found in Labov, Rosenfelder and Fruehwald (2013, 34ff). Unless noted otherwise, the descriptions of the broad workings of FAVE are gleaned from the README files found within the Toolkit. For the purpose of illustration, I use RachelJ’s interview throughout.

3.4.2 FAVE-ALIGN

The FAVE-align toolkit aligns the sound file of the interview with the transcription thereof. Via the HTK Toolkit, each utterance is converted into phonetic notation using the BEEP dictionary, which is then time-aligned with the speech signal (Labov et al. 2013, 35). It splits the text of the transcription into annotation breath groups, which it then feeds into the aligner, the result of which is concatenated into a PRAAT (Boersma & Weenink, 2014) TextGrid file. The first step after the interview process is transcription, for which I used ELAN.³⁸ The transcriber annotates text per breath group (i.e. annotate a section based on when the interviewee takes a natural breath). ELAN (and FAVE) is capable of working with overlapping speech and background noise, both inherent in sociolinguistic interviews. The first few transcriptions were therefore incredibly precise, accounting for every single sound in the entire interview. Though precise and thorough, this process is very laborious, taking around one hour for one minute of conversational speech (especially when the interviewer backchannels unwittingly). In subsequent transcriptions, therefore, only clear, usable utterances were transcribed for each participant. The interviewer's speech was ignored, as were noises (such as laughter or audible breaths) and other sounds that did not overlap with the interviewee's speech.

The transcription is then exported from ELAN as a tab-delimited text file, *RachelJ.txt* (see figure 3.4). FAVE requires the following columns in the text version of the transcription: speaker ID, speaker name, beginning of breath group (in seconds), end of breath group (in seconds), and the transcribed text.

```
RachelJ RachelJ 1209.4 1211.58 But ja, it was really nice.
RachelJ RachelJ 1212.66 1213.53 {LG}
RachelJ RachelJ 1213.53 1218.9 I went to England like absolutely hating beer and now I just drink it all the time. {LG}
RachelJ RachelJ 1221.14 1221.81 {BR}
RachelJ RachelJ 1221.82 1223.86 Like I don't know what it is it's just,
RachelJ RachelJ 1223.86 1224.14 {BR}
RachelJ RachelJ 1224.15 1227.79 I think it's the weather, and the whole pub culture that just makes you,
```

Figure 3.4: Sample of a transcription exported from ELAN as a tab-delimited file.

³⁸ The Language Archive: tla.mpi.nl/tools/tla-tools/elan/. The Max Planck Institute for Psycholinguistics.

Using the Terminal (command-line interface), FAVE-Align runs through *RachelJ.txt* and checks for errors, which include spelling mistakes and errors in transcription conventions (such as indicating laughter using (LG) instead of {LG}). FAVE-Align also identifies words in the transcription that are not present in the dictionary. It saves this information in a text file, *unknownRachelJ.txt*, then opened in Excel (figure 3.5).³⁹

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ABERDEENSHIRE			Aberdeenshire not Aberdeen. {LG}									
2	PROB-	P R AA1 B	PROBABLY	even to get to other places I'd still go via London. So I, I prob- +probably I've been there,									
3	S-	S,SH		It was s- quite like that at some parts but the summer was so hot. {BR} I found it like,									
4	THI-	TH IH1,TH ERD	THINK	They, they did cope okay. Um I thi- +think they got, out of all the kids they the ones that obviously cried, the most. Um,									
5	OBVIOUS-	AA1 B V IY1 Y AH0 S		Um but there were obvious- there were some parts where I just thought, (()) "W-why am I reading {LG} this?" {LG}									
6	REAL-	R IY1 AH0 L	REALLY	'Cause you don't, real- +really like you always wanna go somewhere else, um.									
7	STRA-	S T R EY1	STRANGE	You know cool or cold. {LG} Stra- +strange weird.									
8	T-	T		Other t- other books I've just, there's some parts where I just wanna stop {LG}.									
9	CAPET-	RELEF T	CAPETONIAN	He's a Capet- +Capetonian ja both my parents are from Cape Town.									
10	D-	D	DUNNO	Ja just it's all so green. {LG} I d- +dunno it's so amazing.									
11	Y-	Y,W AY1,Y ERD		And I {LG} always just say like "No, I'm {BR}, I'm not gonna be here {LG} for for the rest of the y-"									
12	PAR-	P AE1 R,P EH1 R,P AH1 R		it seems so huge, and all the par- places are so far apart.									
13	F-	F	FROM	So I've done the whole like th- (()) {LG} and apparently ja, apparently people go f- +from to U C T from Westerford so {LG}. Jal {LG}									
14	J-	JH,Y	JA	English sounding. Ja. J- +ja even though I spent all my time with an Australian {LG}.									
15	W-	W		the only thing that, did, really affect me sometimes was, how it w- got dark so quickly.									

Figure 3.5: Example of a file of unknown elements in a transcription, created by FAVE-Align.

The first column (A) provides the word or false start that the programme did not recognise (though even recognised false starts are included for checking). The second column (B) is for the ARPAbet transcriptions. The third column (C) contains the words related to the false starts if they were provided in the transcription or if they have been pre-programmed. Column four (D) provides the annotation segment where the error was found. Any spelling or convention mistakes are then corrected in the transcription and subsequently deleted from *unknownRachelJ.txt*. Words and false starts that are not present in the dictionary are programmed into the dictionary. For recognised false starts, various options are provided. If the researcher accepts these suggestions, they are deleted from the file. Each unknown word and false start (such as Aberdeenshire and Capet-) is transcribed in column B using the ARPAbet convention, and the file is then resaved as *inputRachelJ.txt*, after the deletion of the third and fourth columns, as is shown in figure 3.6.

³⁹ The user has complete control over the naming of the files. I chose to label files in the way the user manual suggests.

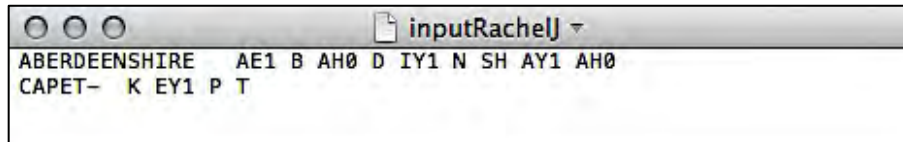


Figure 3.6: Example of an input file to add unknown elements in a transcription to the dictionary.

This input file is then used to recheck the transcription to ensure that all errors were identified and corrected. FAVE-Align saves the results of this second check in a file called *stillunknownRachelJ.txt*. This file should only contain false starts and their respective ARPAbet transcriptions. All words added to the dictionary via the input file are not reproduced here. Instead, they occur in a separate file called *added_dictionary_entries.txt*, from where they are copied over into the master dictionary file, *dict*. In this way, the researcher can make permanent changes to the dictionary, which is not possible in the online version. When this checking step is complete, FAVE-Align uses the *inputRachelJ.txt* file to align the transcription file, *RachelJ.txt*, to the sound file, *RachelJ.wav*, and so produces *RachelJ.TextGrid* (figure 3.7). This takes approximately one hour for an hour-long interview. The TextGrid is opened in PRAAT (along with the corresponding sound file) and includes a phonemic transcription tier (based on the ARPAbet) and an orthographic tier for each speaker transcribed in the interview.

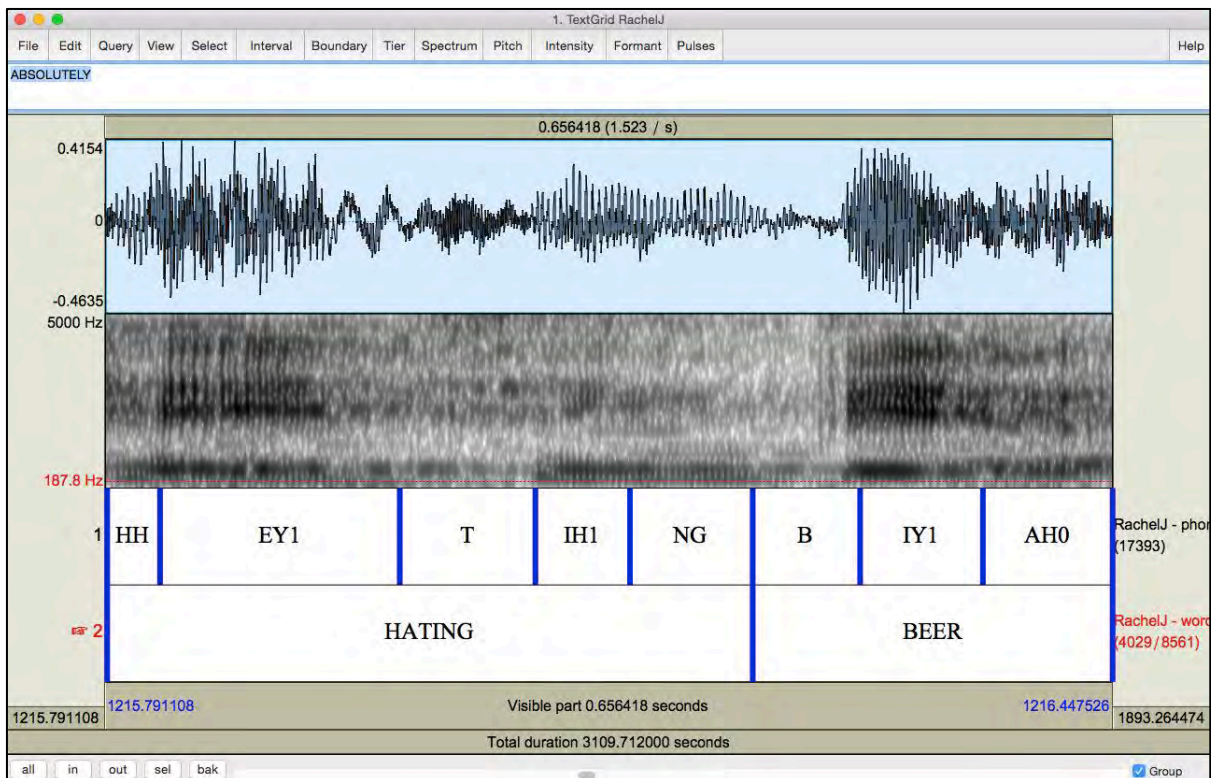


Figure 3.7: Selection of a TextGrid created by FAVE Align.

The TextGrid for each speaker was checked to ensure accuracy of the alignment, particularly given the fact that it had not been used on SAfE data before. Toefy (2014, 99ff) reports a 5% error rate for her alignment process, which largely holds true for my data. I checked the first couple of minutes, a section from the middle, and a section towards the end. Where a TextGrid contained many alignment errors, I went back to the transcription, and often there were bad annotations in it that led to misalignments. A bad annotation in ELAN would be over (or under) selection of the section of the sound wave to be annotated. In cases like these I re-ran the FAVE-Align process with corrected annotations. In most cases, the misalignments occurred in the word list and minimal pairs section of the interview, sometimes due to bad annotations in ELAN. In other cases, as often happened, a word was cut off too early, especially words with final stops (see figure 3.8). In cases like these, the alignments were manually adjusted in the TextGrid (see figure 3.9).

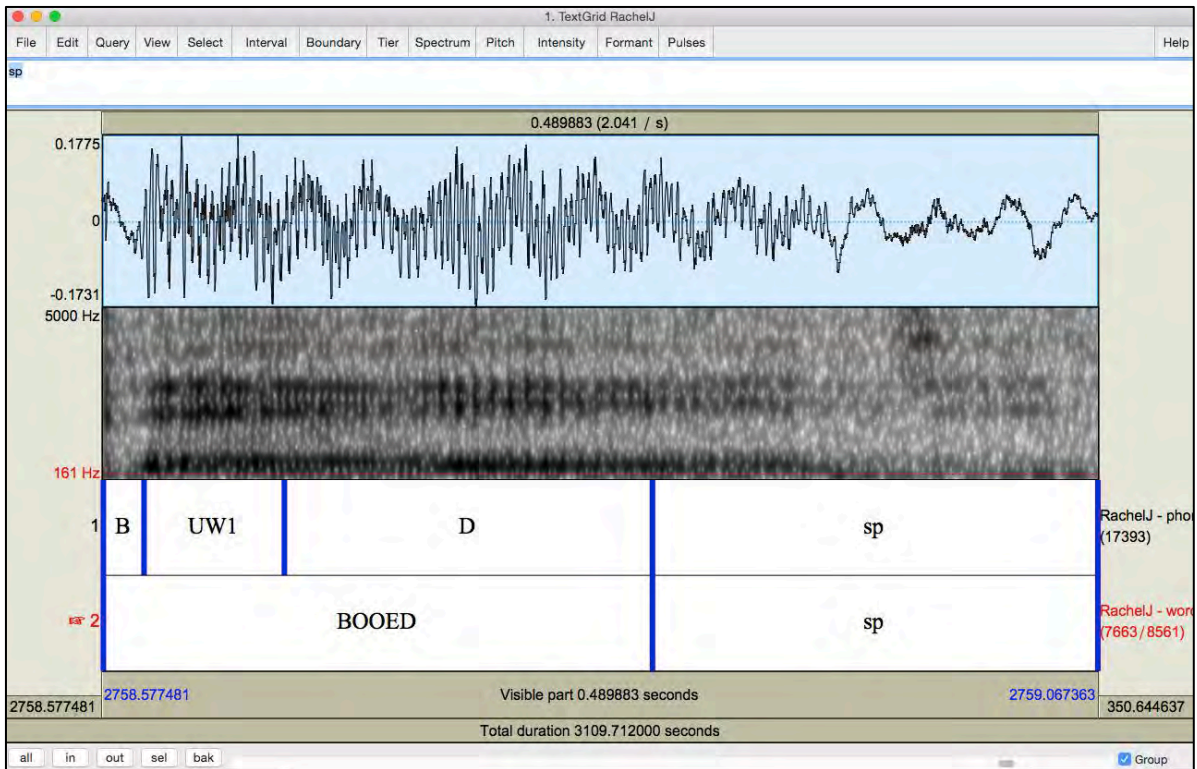


Figure 3.8: Example of a misaligned TextGrid for the word *booed* /bu:d/.

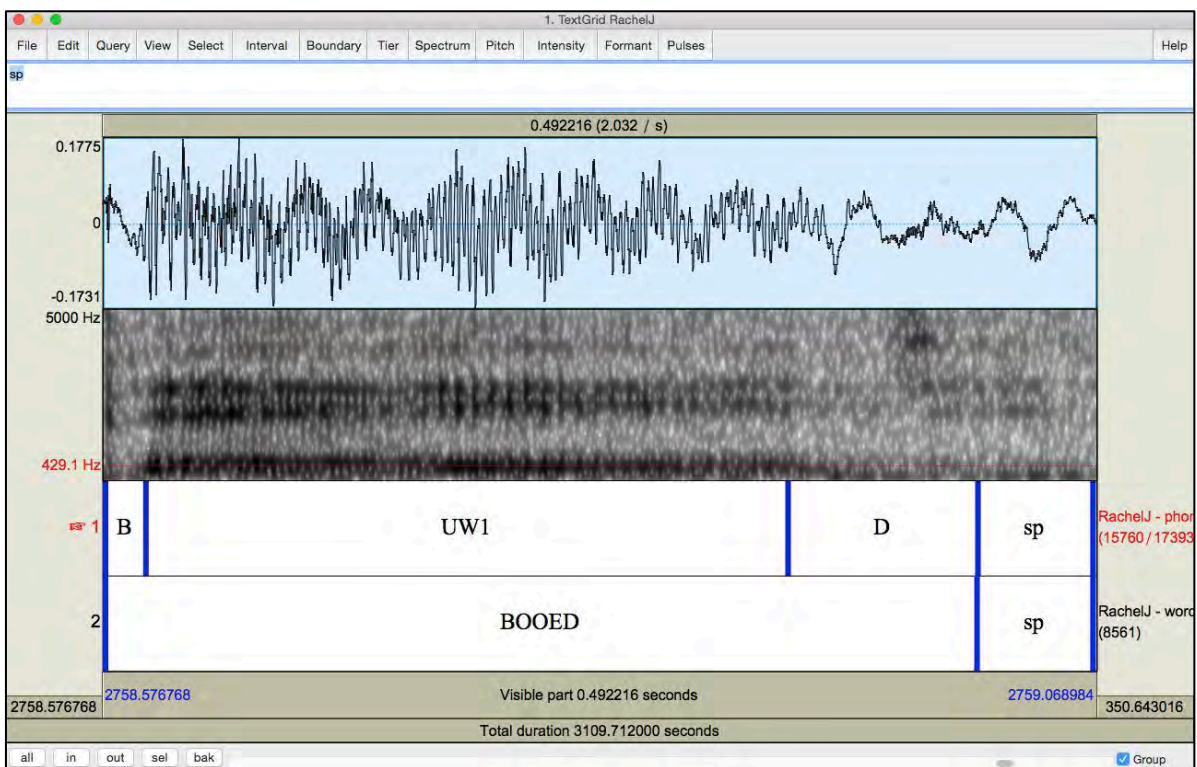


Figure 3.9: Example of a corrected TextGrid alignment for the word *booed* /bu:d/.

Most of the TextGrid alignments for causal speech had an acceptably low error rate (around 5%). The minimal pairs alignments were checked thoroughly given their propensity for misalignment. In general, it would appear that the alignments for interviews that were transcribed very fastidiously (i.e. with numerous speaker overlaps and noise overlap) were those with the most alignment errors. The interviews where only the usable portions of the interviewee's speech were transcribed were far more accurately aligned. Nevertheless, the word list and minimal pairs section of the interview produced the most problems regarding inaccurate alignment. In many, though not all, cases, these misalignments were corrected by correcting the annotation parameters within the ELAN transcription. Corrected TextGrids are then ready for formant extraction via FAVE-Extract.

3.4.3 FAVE-EXTRACT

The FAVE-Extract toolkit is responsible for automatic vowel formant measurement. It uses the TextGrid with its corresponding sound file, extracting the formants for each suitable token. There are various configurations that the user can stipulate for the extraction of formants. These are stipulated (where the desired setting is different to the default setting) in a file called *config.txt*. These can also be specified in the command-line code.

formantPredictionMethod=mahalanobis

Within FAVE-Extract there is a choice between Linear Predictive Coding (LPC) in PRAAT or the Mahalanobis Distance Method for formant prediction. The PRAAT method relies on LPC analysis, where analysed formants are represented by red dots in the wave-form (see figure 3.10). When analysing vowel data manually in PRAAT, the researcher continually changes the number of formants that the LPC analysis searches for in order to optimise the accuracy of the formant tracking. In this sense, the formant tracking for one vowel might be more accurate with five formants, whereas another might require four or six. In figure 3.10 it searched for five formants, which produced accurate formant tracking for F1, F2 and F3.

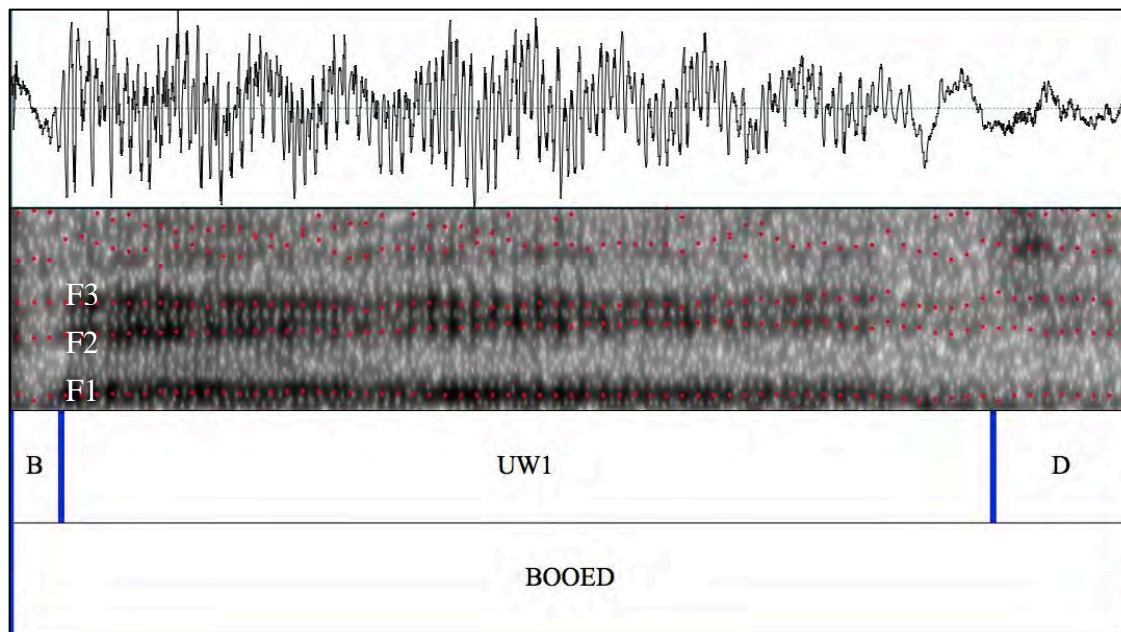


Figure 3.10: Example of PRAAT's formant tracking using LPC analysis.

The default setting for FAVE-Extract is to use these formant readings provided by PRAAT. However, continual adjustment is necessary to ensure the accuracy of the formant tracking and therefore the formant analysis (see Toefy 2014, 105-6). In this regard, the Mahalanobis method for formant prediction is the preferred choice. Evanini (2009) developed this algorithm, which “compares all poles and bandwidths returned by the LPC analysis for the vowel to a distribution of expected formant poles and bandwidths taken from the ANAE⁴⁰ measurements” (FAVE-Extract README). Essentially, this method simulates the continual adjustments that a researcher would carry out in a manual analysis, and provides much more accurate formant readings (Toefy 2014, 106). When using the Mahalanobis method, *remeasurement=T* tells FAVE to analyse the formants twice, the second time by using the individual speaker's vowel system as the base of comparison for the distance algorithm. The final formant measurement chosen per token is that with the smallest Mahalanobis distance (Labov et al. 2013, 36).

⁴⁰ Atlas of North American English (Labov, Ash & Boberg, 2006).

measurementPointMethod=faav

Once the most accurate method of formant tracking has been identified, a decision regarding the exact point in the duration of the vowel at which a formant should be measured must be made. FAVE-Extract provides numerous options, of which the default ‘faav’ was chosen. After testing various methods, Evanini (2009, 60-6) ascertained that measurement a third of the way into the vowel produced the best automatic formant readings. The faav method measures the vowel one third of the way through its duration, with the following modifications (Labov et al. 2013, 35):

1. /aɪ/ and /eɪ/ are measured at the maximum F1.
2. /oo/ and /aʊ/ are measured halfway between vowel onset and the maximum F1.
3. /u:/ after coronal consonants is measured at the beginning of the vowel.

In order to avoid the measurement of reduced tokens, *removeStopWords=T* instructs FAVE-Extract to skip over stop words (such as *the* and *and*). Only vowels with 0.05 seconds or longer are measured to further avoid unstressed tokens, though the researcher can specify the vowel length to be considered.

With the set configurations, FAVE-Extract thus runs through the TextGrid and sound file of each speaker, and extracts the formant information of each appropriate token in the interview. For each speaker, it provides two separate output files: one containing raw formants (*RachelJ.txt*) and one with normalised formants (*RachelJ_norm.txt*). The normalised formants are used for the analyses in the following chapters. These are comparable across speakers, since any physiological differences (based on age or gender) are adjusted for. The default normalisation method is Lobanov (1971), yielding a z-score for each token which is then rescaled to hertz-like values (Labov et al. 2013, 36).

Labov et al. (2013, 37-8) showed that FAVE-produced formant readings are highly accurate for the Philadelphia Neighbourhood Corpus. As will be shown in Chapters 4 and 5, the formant analyses via FAVE work for SAfE and are thus accepted as accurate. However, the process requires more manual corrections and realignments than it does for American Varieties of English, which adds time to the task.

3.4.4 WORKING WITH THE OUTPUT

The normalised output of each speaker is opened in Excel, where it is coded for the various factors pertinent to the study. However, the data produced by the FAVE-Extract process must be carefully considered, given the various differences between the American varieties of English and SAfE. This involves thorough correction of vowel assignment so that they meet the following criteria:

1. AA refers to BATH /ɑ:/ alone and does not include LOT /ɒ/. All LOT tokens are assigned the symbol OH.⁴¹
2. STRUT /ʌ/ tokens are distinguished from schwa /ə/ tokens. The BEEP dictionary codes both as AH0. Tokens analysed as AH that were schwa tokens were removed.

Other less systematic errors also had to be corrected. A prevalent example is the word *does* /dʌz/, which was regularly analysed as /doʊz/. After the correction of vowel assignment, the tokens were worked through on an individual basis. Any gross errors in formant measurement by FAVE-Extract, identified by inspection, were removed. All words which were not to be included in the analyses were also removed. These were various words of non-English origin (such as *Afrikaans* and *vloek*), or regular discourse markers (such as *like* and *ja*).⁴² Readings for vowels that did not carry primary stress (or those likely to be reduced, such as *can*) were also removed. In total, almost 39,000 tokens were removed leaving 100,848 tokens suitable for analysis.

⁴¹ This step occurs at the output stage given the complexity (or impossibility) of adding this OH phoneme to the FAVE Toolkit.

⁴² *Vloek* /fløk/ is an Afrikaans verb meaning ‘to swear’ which is regularly used in English. *Ja* is a discourse marker prevalent South Africa, similar to ‘yeah’ in other varieties of English.

3.5 STATISTICAL APPROACHES AND GRAPHICAL PRESENTATION

The formant values that provide information on vowel height (formant 1) and frontness (formant 2) form the basis of the analyses presented in this thesis. The statistical analyses of these values were undertaken via the open source statistical programme R (R Core Team 2014). In particular, linear mixed-effects modelling, random forests, conditional inference trees, Pillai Scores and Welch's Two Sample *t*-tests were used. I discuss the particulars of each in section 3.5.1, drawing extensively from the descriptions provided by Hall-Lew (2010), Tagliamonte (2012), Tagliamonte and Baayen (2012) and Strelluf (2014). Thereafter I outline various ways in which I present the data visually.

Prior to statistical analysis and visual presentation, the data was sorted and coded for various factors. Each vowel token in the dataset carries the following information:

1. Speaker: the pseudonym of the participant.
2. Vowel: the ARPAbet notation appropriate for each vowel.
3. Word: the word containing the vowel in question.
4. F1: the reading of formant 1 in normalised hertz-like values.
5. F2: the reading of formant 2 in normalised hertz-like values.
6. Style: the portion of the interview from whence the word originates (CS = casual style, MP = minimal pairs test).
7. Sex: the gender of the participant (G = female, M = male).⁴³
8. Ethnicity: the ethnicity of the participant (W = white, B = black).
9. Environment: the phonetic environment of the vowel in question. Detailed descriptions of the environments for each vowel are given in Chapters 4 and 5.
10. Date of birth: the participant's year of birth. I chose not to group the participants into age brackets so as to maximise the ability of the statistical modelling to elucidate any age effects.

The coding of the data was carried out in Excel for MAC (2011). It was then exported as a tab-delimited text file to be imported into R for statistical modelling and graphical production.

⁴³ I chose to use 'G' for 'female' instead of 'F' to avoid potential confusion between 'F' for 'female' and 'F' for 'formant'.

3.5.1 STATISTICAL TESTING IN R

As a general, first step in exploring the data statistically, linear mixed-effects regression (lmer) was used, a method introduced to sociolinguistics by Johnson (2010). This kind of statistical modelling is particularly well suited to sociolinguistic data, which is often unevenly distributed across individuals, *inter alia*. Mixed-effects models are very powerful and principled, successfully dealing with the different kinds of predictors inherent to sociolinguistic data (Tagliamonte & Baayen 2012, 142-3). The model includes two different effect factors (predictors): fixed-effect predictors and random-effect predictors. ‘Sex’, ‘date of birth’, ‘phonetic environment’ and ‘ethnicity’ are all fixed-effect predictors in that their representation in the data is fixed. In other words, for ‘environment’ there are a number of categories to which the data can adhere (as set out by the researcher), and all the categories are represented by the observations in the dataset. Similarly, ‘ethnicity’ and ‘sex’ are limited to either ‘black’ or ‘white’; and ‘male’ or ‘female’, and the data represents all the options available.

Random-effect predictors are those of which only a subset is represented in the data. For example, the individuals comprising the dataset are not representative of all possible Capetonian individuals; they are but a subset. Similarly, casual style data consists of words per individual that are not uniform or fixed: the researcher has no control over which words a participant will use, and some participants will contribute more words than others.⁴⁴ In addition, individuals in a sociolinguistic study provide numerous observations, and as such constitute sources of variation that the mixed-effects model accounts for. This is particularly important, since if a random-effect predictor is analysed as a fixed-effect predictor, the analyses and conclusions based on the statistical modelling will apply only to the individuals within the dataset. If random-effect predictors are analysed as such, the conclusions drawn of the analyses may be generalised as applicable to or characteristic of the greater population of which the dataset is a sample (Tagliamonte & Baayen 2012, 143).

The R-package used for linear mixed-effects regression is {lmerTest} (Kuznetsova, Brockhoof & Haubo, 2014). This package is very similar to another well-known package, {lme4} (Bates, Maechler, Bolker & Walker, 2014), except that in the analysis of variance

⁴⁴ In minimal pairs testing, the researcher has complete control over which word the participant uses, and each participant contributes the same amount of words. In these tests, therefore, ‘word’ would not be included as a random-effect.

(ANOVA) function, p -values are calculated from f -statistics using “Satterthwaite” approximations. Furthermore, the summary function provides p -values and degrees of freedom for the t -tests. Table 3.3 provides an example of the coding and output of the `lmer` in R, testing the extent to which KIT frontness (F2) is predicted by the various fixed-effects.

The `lmer` function measures and compares each factor to an intercept, explained by Tagliamonte (2012, 148) as follows:

R uses “treatment” coding for the assessment of factor groups. In treatment coding one of the factors in each factor group is selected as the baseline or reference level (intercept). This means that a value is provided only for the factors that are not the reference level; the values reflect the contrast between the reference level and the other levels. As a default, R will select the factor whose label comes first alphabetically as the reference level.

I coded my data in a way that presented the free environments first (as zero) to which each conditioning environment is then compared.⁴⁵ In the summary(`kit.lmerf2`) results in table 3.3, therefore, the intercept is ‘KIT0’ for environment, ‘1954’ for date of birth, ‘black’ for ethnicity and ‘female’ for gender. The researcher however has the option to change the intercept within R at any point during the analysis.

When compared to the intercept, each of the remaining fixed-effect factors (except KIT3) is significantly different according to the p -value scores. The summary of (`kit.lmerf2`) provides a plethora of information regarding the datasets tested (see Tagliamonte 2012, 147ff).⁴⁶ For the purposes of this study, I refer to and quote the estimates as a way of obtaining vowel means (as per Strelluf 2014, 87) and the p -values to gauge significance (measured at $p \leq 0.05$).⁴⁷

⁴⁵ Detailed descriptions of the environments I coded my data for are provided in Chapters 4 and 5.

⁴⁶ Tagliamonte (2012) used an older version of the {lme4} package, and as such her descriptions do not account for the output of {lmerTest} in its entirety. Nevertheless, it is a very helpful description.

⁴⁷ To obtain accurate formant means, only one fixed-effect factor is used at a time, i.e.:
`kit.lmerf2<-lmer(f2 ~ env + (1|speaker) + (1|word), data = kit)`

```

> kit.lmerf2<-lmer(f2~env+dob+ethnicity+sex+(1|speaker)+(1|word), data =
all[all$vowel=="IH",])
> anova(kit.lmerf2)
Analysis of Variance Table of type 3 with Satterthwaite
approximation for degrees of freedom

```

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
env	6050138	1512535	4	1127.07	32.257	< 2.2e-16 ***
dob	242855	242855	1	54.13	5.179	0.02684 *
ethnicity	268288	268288	1	42.96	5.722	0.02120 *
sex	1705108	1705108	1	45.28	36.364	2.756e-07 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(kit.lmerf2)
Linear mixed model fit by REML t-tests use Satterthwaite approximations
to degrees of freedom [merModLmerTest]
Formula: f2 ~ env + dob + ethnicity + sex + (1 | speaker) + (1 | word)
Data: all[all$vowel == "IH", ]

REML criterion at convergence: 95252.5

Scaled residuals:
  Min       1Q   Median       3Q      Max
-4.0588 -0.6024 -0.0445  0.5826  4.9873

Random effects:
 Groups   Name      Variance Std.Dev.
word     (Intercept) 11149    105.59
speaker  (Intercept)  2268     47.63
Residual                    46889    216.54
Number of obs: 6974, groups: word, 702; speaker, 53

Fixed effects:

```

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	5622.7680	1634.0734	54.1000	3.441	0.00112 **
envkit2	83.4388	13.2212	1483.1000	6.311	3.65e-10 ***
envkit3	-55.0755	29.7888	1194.1000	-1.849	0.06472 .
envkit5	-61.8407	15.7968	1202.8000	-3.915	9.56e-05 ***
envkit6	-63.4339	27.2904	953.9000	-2.324	0.02031 *
dob	-1.8664	0.8201	54.1000	-2.276	0.02684 *
ethnicityW	-45.9981	19.2299	43.0000	-2.392	0.02120 *
sexM	87.0817	14.4407	45.3000	6.030	2.76e-07 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Table 3.3: Example of R codes and results for a linear mixed-effects regression model.

The interactions between certain fixed-effects may also be tested via the lmer model, using a * in the code, presented in table 3.4. This code asks R to use linear mixed-effects regression to analyse the extent to which ‘environment’ predicts the height of KIT, and in addition, to test the interaction between ‘environment’ and ‘date of birth’.⁴⁸

```

> lmerf1<-lmer(f1~env*dob+(1|speaker)+(1|word),data =all[all$vowel=="IH",]
> summary(lmerf1)
Linear mixed model fit by REML ['merModLmerTest']
Formula: f1 ~ env * dob + (1 | speaker) + (1 | word)
Data: all[all$vowel == "IH", ]

REML criterion at convergence: 79194.7

Scaled residuals:
   Min       1Q   Median       3Q      Max
-4.0372 -0.5862 -0.0475  0.5009 10.3989

Random effects:
 Groups   Name                Variance Std.Dev.
 word    (Intercept)           635.3    25.21
 speaker (Intercept)           279.0    16.70
 Residual                            4751.2    68.93
Number of obs: 6974, groups: word, 702; speaker, 53

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  1058.8244   689.7830   145.0000    1.535 0.126965
envkit2      -2002.6615   545.7045  6595.0000   -3.670 0.000245 ***
envkit3      -1620.9667  1258.9185  6535.0000   -1.288 0.197935
envkit5       -192.6635   744.8288  6810.0000   -0.259 0.795899
envkit6      -1823.1222   922.3467  6778.0000   -1.977 0.048126 *
dob           -0.2591     0.3469   145.0000   -0.747 0.456364
envkit2:dob    1.0086     0.2744  6594.0000    3.676 0.000239 ***
envkit3:dob    0.8078     0.6332  6536.0000    1.276 0.202101
envkit5:dob    0.1010     0.3744  6808.0000    0.270 0.787279
envkit6:dob    0.9197     0.4639  6776.0000    1.983 0.047446 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Table 3.4: Example of R codes and results for a linear mixed-effects regression model, including factor interactions.

The results in the summary show that KIT2 and KIT6 are significant predictors of F1 ($p=0.000$ and $p=0.048$ respectively), and that there is significant interaction between KIT2 and ‘date of birth’ ($p=0.000$), as well as KIT6 and ‘date of birth’ ($p=0.047$). Based on these analyses, the interaction between the environments of KIT and the age of the participants is a significant area to be explored in detail, which would not have been apparent in the earlier tests (see Tagliamonte & Baayen 2012, 152ff). In both cases, the exact nature of the statistical difference is however difficult to pinpoint.

⁴⁸ Chapter 4, section 4.2 provides a description of the different environments of KIT.

The negative side of linear mixed-effect regression is that it does not effectively manage distortion caused by outliers (Tagliamonte & Baayen 2012, 144). There are various ways of excluding potential outliers from the vowel data that could minimise this problem (see Thomas 2011, 159). Since this is a study on vowel chain shifting, ‘advanced shifters’ could for example be identified as outliers compared to the other speakers in the sample. Potential outliers could also be as a result of mismeasurement by FAVE-Extract, though the careful checking process would likely have removed the bulk of these. Following Thomas (2011, 158) therefore, potential outliers were not excluded from the data, which necessarily impedes the efficacy of the lmer models. Moreover, that the model fits and accurately describes the data is something one can often not be completely sure of (see Tagliamonte & Baayen 2012). It is for this reason that additional statistical testing was applied to the data. The lmer models were used to tease out general patterns to guide further statistical analysis, subsequently carried out with random forests and conditional inference trees (ctrees) for vowel chain shift analyses.

Tagliamonte and Baayen (2012) describe the use of random forests (and therefore ctree) in addition to linear modelling as highly advantageous to sociolinguistic research, particularly in that they provide nuanced understanding of the phenomena under investigation. Random forests and conditional inference trees are processed in R using the `{party}` package via the `cforest` function (Strobl, Boulesteix, Kneib, Augustin & Zeileis 2008; Strobl, Boulesteix, Zeileis & Hothorn 2007; Hothorn, Buehlmann, Dudoit, Molinaro & Van Der Laan 2006). Conditional inference trees are also generated in `{party}`, but with the `ctree` function (Hothorn, Hornik & Zeileis, 2006). The R package `{partykit}` is a similar alternative for these kinds of statistical modelling (Hothorn & Zeileis 2014).

Random forests, like lmer models, analyse the extent to which a variable is predicted by the various factor groups (predictors) within a dataset. The `cforest` function instructs R to work through the data painstakingly, and by trial and error, to identify whether or not the factor group predicts the variable (F1 or F2 in this case). Tagliamonte (2012, 153) notes that this kind of analysis is “excellent for exposing the relative contribution (i.e. strength) of each factor group on the variable under investigation.” During this process, the `cforest` function creates many conditional inference trees, which amount to the random forest. For a detailed description of the inner workings of this model, see Tagliamonte and Baayen (2012, 159ff). Table 3.5 presents the code and output for the `cforest` function in R.

```

> kit.cforest=cforest(f2~env+dob+ethnicity+sex,data=all[all$vowel=="IH",])
> kit.cforest.varimp=varimp(kit.cforest,conditional=TRUE)
> dotplot(sort(kit.cforest.varimp))

```

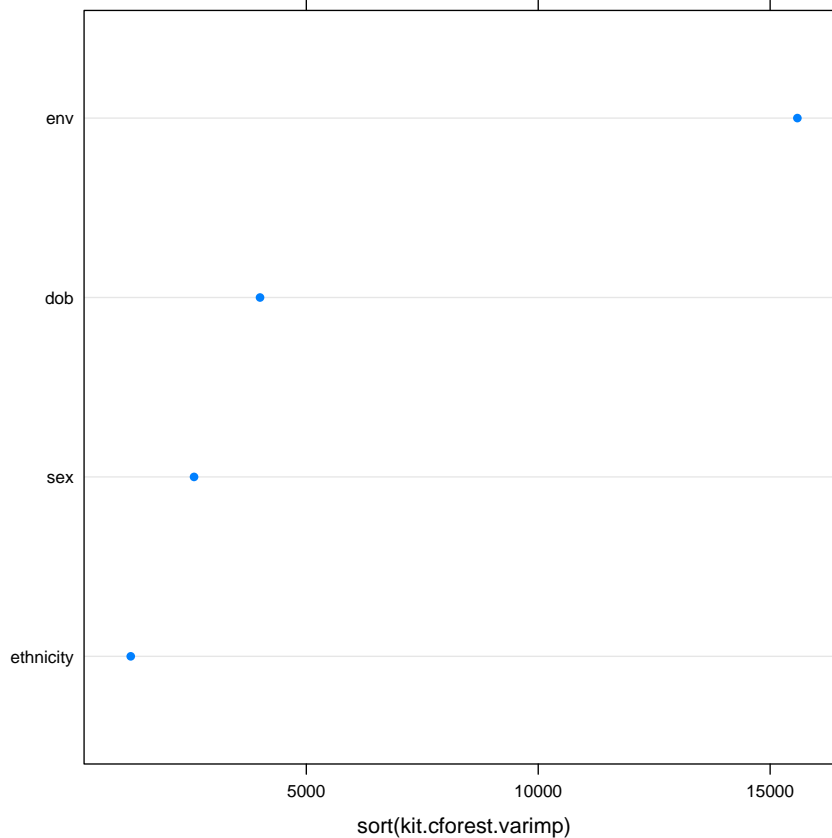


Table 3.5: Example of R codes and results for a random forest model.

The plot of the cforest (table 3.5) indicates that the frontness (F2) of the KIT vowel is most strongly predicted by phonetic environment, followed at a distance by date of birth and gender. Comparatively, ethnicity does not contribute to the prediction of KIT F2 in as influential a way. Each of these fixed-effects was identified as significant predictors in the lmer model (see table 3.2), though the random forest has elucidated the hierarchical importance of each. If a predictor is not deemed valuable (i.e. it is not a strong predictor), it can be excluded as a factor in subsequent analyses. This is particularly useful if there are a large number of predictors within a dataset, though generation of the random forests with many predictors is a lengthy enterprise (Tagliamonte & Baayen 2012, 165). There are only four predictors in my dataset, so each random forest model took about 10 minutes to compute.

The information provided by the random forest is essential for the analysis of each vowel involved in a possible chain shift, though it does not specify *how* the predictors impact the variable. Furthermore, the random forest does not provide any indication of how the predictors might work together. In order to gain insight into these areas, a conditional inference tree is used, which “is able to depict the subtle interactions in the data using a hierarchical display” (Tagliamonte 2012, 153). According to Horthorn et al. (2015, 8), this is achieved “by binary recursive partitioning in a conditional inference framework.” In this sense, it presents the interaction of the factor groups as a tree with binary branches where statistically applicable. It also provides the *p*-values upon which this branching is based. Table 3.6 presents the R code for the *ctree* function, as well as an example of a *ctree*. *Ctrees* can become very large and complicated. For ease of presentation here, the example includes only ‘environment’ and ‘date of birth’ as the most valuable predictors identified by the random forest model.

The predictors and their effect on the variable are ranked hierarchically. The *ctree* in table 3.6 therefore shows that ‘environment’ is a strong, statistically significant predictor of the F2 of KIT. Hierarchically it is in position 1, 2 and 3 after various binary partitioning. ‘Date of birth’ is ranked lower down in positions 7 and 8. This mirrors the results of the random forest model (table 3.5), which showed a high value for ‘environment’, and a lower value for ‘date of birth’. The *ctree* also shows that ‘date of birth’ interacts with ‘environment’, but only for the environment KIT2, where there are significant age differences.⁴⁹ In this way, the *ctree* informs the researcher of the hierarchy in which the variables are predicted by the various factors, and how the factors relate to one another.

⁴⁹ This environment includes KIT where it occurs before velars, word initially and after /h/.

```

> kit.ctree=ctree(f2~env+dob,data=all[all$vowel=="IH",])
> plot(kit.ctree)

```

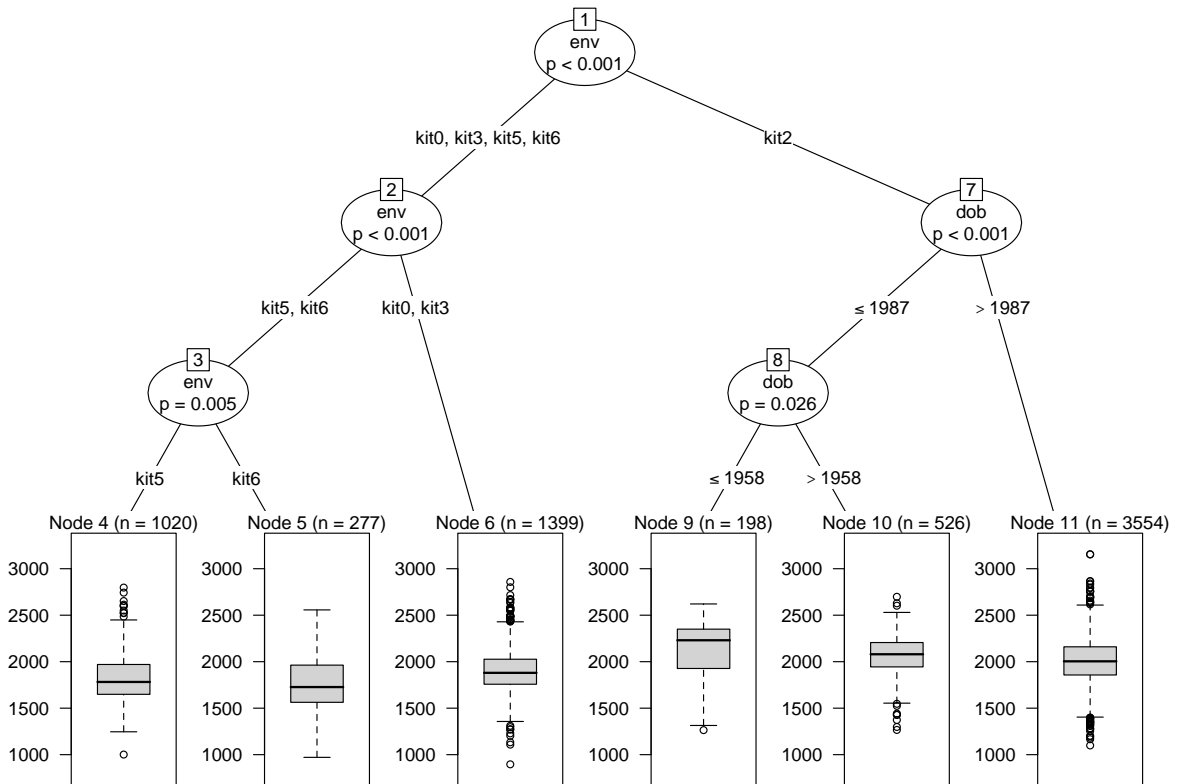


Table 3.6: Example of R codes and results for a conditional inference tree.

In addition, the `ctree` informs the researcher of exactly how the factor groups predict or influence the variable. Vowel formant data are presented via boxplots (also called box-and-whisker plots) once the recursive binary branching has reached its end (nodes 4-6; 9-11 in table 3.6). Boxplots provide fairly detailed information about the spread of the data that it represents. Figure 3.11 presents node 11 of the `ctree` from table 3.6. Each point, indicated by arrows, provides specific information.

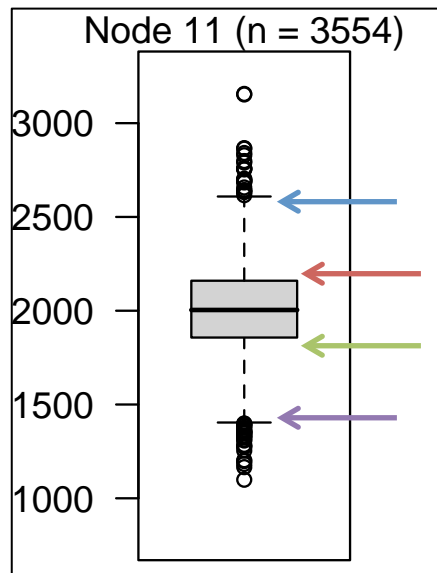


Figure 3.11: Example boxplot for the F2 of KIT2 (node 11, table 3.5).

This boxplot tells us that, for KIT2 for speakers born after 1987, there are 3554 tokens in total. The spread of the tokens are represented as follows (Toefy 2014, 112): the blue arrow points to a line indicating the maximum F2 value in the dataset. The red arrow points to the uppermost line of the box, joined to the maximum value, barring outliers, by a whisker. This represents the upper quartile of the data. Another whisker joins the lowermost line of the box (green arrow) and the minimum F2 value within the dataset (purple arrow). This represents the lower most quartile of the data: 25% of the data are lower in value than the lowermost line of the box. The black line within the box indicates the median of the data: 50% of the tokens are higher than this value, and 50% are lower. The circles above and below each whisker indicate the outliers present in the data. The outliers are identified as tokens with values 1.5 times the range above the upper quartile and the lower quartile. The boxes therefore represent the middle 50% of the data. The boxplots contained in each ctree output are then compared. For F2, higher boxes indicate frontier vowels. For F1, higher boxes indicate lower vowels.

In addition to the lmer, cforest and ctree functions for statistical analysis, three further tests and measures were employed to analyse vowel merger. Following Strelluf (2014), Euclidean Distance measures, Welch’s Two Sample *t*-tests and Pillai Scores were run in R. Each of these tests calculates the distance between vowels and assigns a significance value to the distance. The Euclidean Distance “measure(s) a straight line distance between the means of two vowels as plotted on a two-dimensional space defined by F1 and F2” (Strelluf 2014, 91). It calculates the distance by using the formula below, followed by the R-code for which a

specific package is not required. The output in R provides a single measure of distance in hertz between the mean (both F1 and F2) of TRAP and STRUT for each speaker.

$$\sqrt{(\text{trapf1mean}-\text{strutf1mean})^2 + (\text{trapf2mean}-\text{strutf2mean})^2}$$

```
> sqrt(((trapag[3]-strutag[3])^2)+((trapag[4]-strutag[4])^2))
```

The distance between tokens can also be tested via Welch's Two-Sample *t*-tests in R, as shown in table 3.7:

```
> chloetrap=trap[trap$speaker=="ChloeV",]
> chloestrut=strut[strut$speaker=="ChloeV",]
> x=chloetrap[4]
> y=chloestrut[4]
> t.test(x,y)

Welch Two Sample t-test

data: x and y
t = 4.5175, df = 125.158, p-value = 1.429e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 46.93245 120.11673
sample estimates:
mean of x mean of y
 830.0000  746.4754

> x=chloetrap[5]
> y=chloestrut[5]
> t.test(x,y)

Welch Two Sample t-test

data: x and y
t = 5.3825, df = 105.559, p-value = 4.467e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 69.91008 151.45103
sample estimates:
mean of x mean of y
1596.681 1486.000
```

Table 3.7: Example of Welch's Two Sample *t*-test in R for the F1 [4] and F2 [5] of ChloeV's TRAP and STRUT vowels.

A Welch Two-sample *t*-test calculates a *t*-score and a *p*-value for F1 and F2 separately (table 3.7). Evanini (2009) suggests that a Euclidean Distance below 100 Hz is indicative of vowel merger. In this sense, a Euclidean Distance of 100 Hz or less in conjunction with a low *t*-score and non-significant *p*-value (for both F1 and F2) would thus indicate that the speaker (or group, depending on the test) displays vowel merger.

Though these measures and tests are useful in analysing potential vowel merger, they are flawed in various ways. The Euclidean Distance measure relies on mean values, which eliminates any information on variation within an individual's vowel data. Whilst the Welch *t*-test measures every token, it separates F1 from F2, obscuring the holistic picture of potential vowel merger. In addition to these measures and tests, therefore, I also use Pillai Scores. This is a beneficial tool in that it takes phonological environment into account, and deals effectively with the unbalanced nature of sociolinguistic interview data (Hall-Lew 2010, 3). Moreover, this method takes individual tokens into account as well as F1 and F2 as a whole, as explained by Strelluf (2014, 92):

A Pillai score is an output of a multivariate analysis of variation (MANOVA) test. Rather than calculating mean values, the Pillai score takes every data point into account. Given an explanatory factor, it reports how much dispersion exists across data points. More overlap generates a lower Pillai score and more dispersion a higher Pillai score. The MANOVA also returns a *p*-value to estimate whether the samples created by the explanatory factors suggest different populations.

Table 3.8 illustrates the method of obtaining Pillai Scores in R, for which no specific R-package is required. The output provides the results per speaker in alphabetical order (the first three of which are reproduced in table 3.8). In each case, the function provides a Pillai Score and a *p*-value for the vowel as a factor ('vowel') and for the vowel's phonological environment as a factor ('env'). The Pillai Score for 'vowel' indicates to what extent dispersion exists in the data. The Pillai Score (and *p*-value) for 'env' indicates whether or not phonological environment has a significant impact on the variation within the data (Hall-Lew 2014, personal communication). I quote and use the Pillai Score and *p*-value for 'vowel' in most cases, and use ctree to tease out environmental effects on the closeness of two vowels. A Pillai Score close to 1 indicates dispersion in the data, whereas a Pillai Score close to 0 indicates very little dispersion in the data. In this sense, a speaker who has a low Pillai Score combined with a non-significant *p*-value is a speaker showing possible merger.

```

> trapstrutpillai$subj = as.factor(trapstrutpillai$speaker)
> subs=levels(as.factor(trapstrutpillai$speaker))
> for(sub in subs) {
+   tmp<-subset(trapstrutpillai,speaker==sub)
+   f1f2=cbind(tmp[,5],tmp[,6])
+   vowel<-factor(tmp[,2])
+   env<-factor(tmp[,4])
+   fit<-manova(f1f2 ~ vowel + env)
+   print(sub)
+   print(summary(fit, test="Pillai"))}

[1] "AaronA"
      Df Pillai approx F num Df den Df Pr(>F)
vowel  1 0.58533  145.392   2   206 < 2e-16 ***
env    12 0.16254   1.526  24   414 0.05474 .
Residuals 207
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "AlanL"
      Df Pillai approx F num Df den Df Pr(>F)
vowel  1 0.71080  314.605   2   256 < 2.2e-16 ***
env    12 0.22061   2.655  24   514 4.12e-05 ***
Residuals 257
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[1] "BobbyC"
      Df Pillai approx F num Df den Df Pr(>F)
vowel  1 0.44184   34.831   2    88 7.200e-12 ***
env    12 0.53499   2.708  24   178 9.708e-05 ***
Residuals 89

```

Table 3.8: Example of a Pillai Score generation in R for TRAP and STRUT [2] per speaker for F1 [5] and F2 [6].⁵⁰

If in addition to these a speaker also has 100 Hz or less Euclidean Distance between the vowels and a non-significant p -value returned by a Welch t -test, it is likely that the two vowels are very close together. This may mean vowel merger, though Hall-Lew (2010, 5) notes that speakers who have low Pillai Scores together with non-significant p -values could also display near-merger rather than full vowel merger. Using all three tests in this way yields statistically well-corroborated analyses.

⁵⁰ Hall-Lew (2010) provides an excellent description of Pillai Scores and their implementation in R (also <http://www.lel.ed.ac.uk/~lhlew/pillai.html>).

3.5.2 GRAPHICAL PRESENTATIONS OF THE STATISTICAL AND VOWEL DATA

In describing and analysing the vowel data, I present various figures and tables. General visual representations of the vowels are presented as scatterplots, most often based on vowel means per speaker. These plots were generated using the R-package `{lattice}` (Deepayan, 2008). In some cases, particularly concerning vowel merger, standard deviation ellipses (x2) are included to provide a visual indication of the spread and/or overlap of vowels. These were generated on top of the scatterplots using the R-package `{vowels}` (Kendall & Thomas, 2014). The analysis of vowel chain shift and merger will largely make use of conditional inference trees as statistical and visual representations of the data. The random forest analyses are not reproduced in the analysis chapters, since they informed the factor selection in the ctree productions and were thus not the primary source of analysis. The results of the linear mixed-effects models are largely quoted in-text. Pillai Scores, Euclidean Distances and Welch *t*-test results are provided in tables or appendices as appropriate.

CHAPTER 4: RESULTS: SHORT FRONT VOWELS

This chapter presents the detailed acoustic results of the short front vowels KIT, DRESS and TRAP. I begin with an overview of the whole vowel system in section 4.1. Each short front vowel is then dealt with individually from section 4.2 to 4.4. The details of the Reverse Short Front Vowel Shift (henceforth the Reverse Vowel Shift) as present in the speech of white Capetonians are presented after these analyses in section 4.5. The way in which other Capetonians, particularly black speakers, participate in the chain shift is discussed in section 4.6. Where appropriate, brief mention is made of any outliers. The results are presented visually via formant (scatter) plots and conditional inference trees.

4.1 OVERVIEW OF THE VOWEL SPACE

Prior to the investigation into the finer workings of the vowels in question, it is essential to provide an overview of the data. In this vein, the Capetonian English vowel system is presented in figure 4.1, using mean values per speaker. Mean values are not maximally efficient at illuminating inherent or allophonic variation, though they do provide an overview of the vowel system for descriptive purposes. Generally, the following well-attested features of SAfE occur in the data (as discussed in Chapter 2):

- a) Fronted GOOSE.
- b) Fronting FOOT.
- c) Centralised KIT, relative to FLEECE.
- d) Centralised DRESS, compared to KIT.
- e) Lowered, retracting TRAP.
- f) Fronted (or non-back) LOT.
- g) High, front FLEECE.
- h) BATH range showing backing and raising.
- i) Raised THOUGHT.

Given the positioning of LOT and THOUGHT, Low Back Merger is not a feature of SAfE as it is in some American and Canadian varieties of English. In maintaining a difference between LOT and THOUGHT, SAfE is similar to southern varieties of American English as well as English in New Zealand, Australia and the United Kingdom.

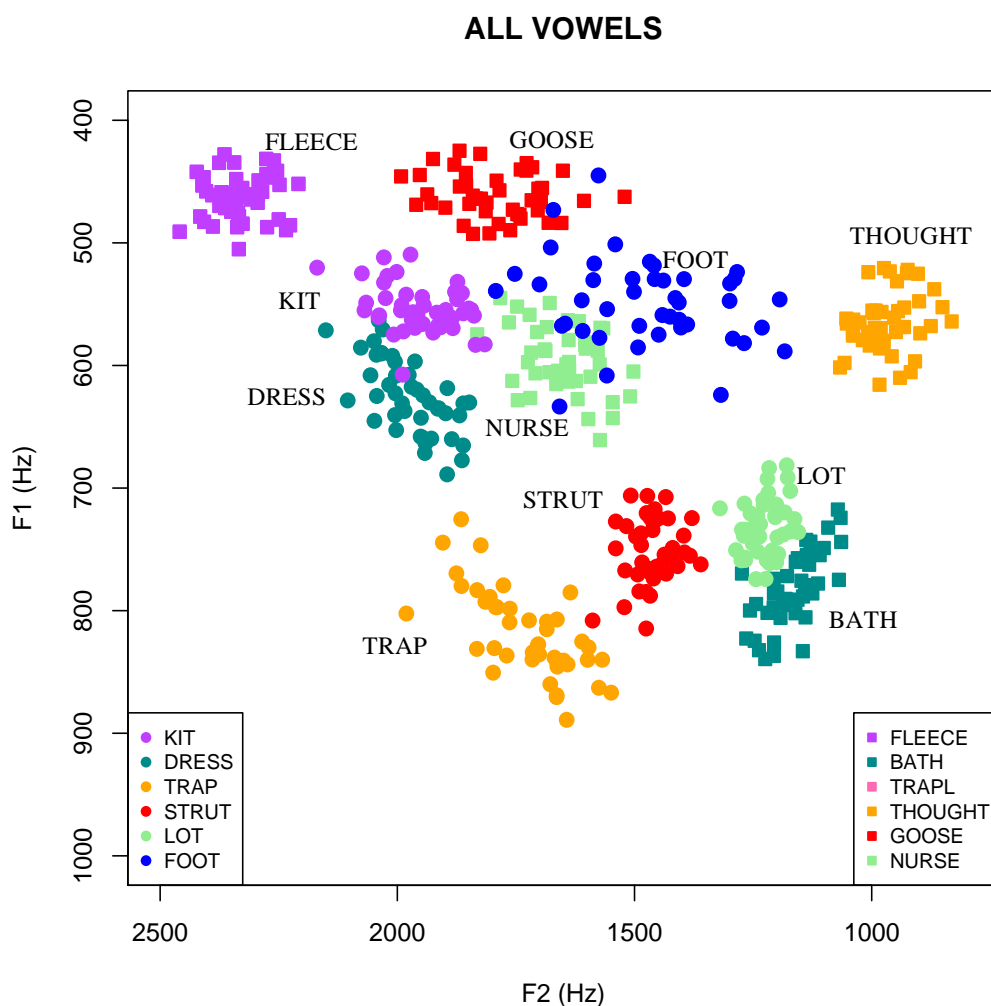


Figure 4.1: Means per speaker of each monophthong (in all environments) in the sample of Capetonian English speakers.

Figure 4.2 presents the same set of vowels, separated by age. ‘Older’ speakers are classified as those born between 1954 and 1961, and ‘younger’ speakers are those born between 1981 and 1995. At the time of the interviews (mostly done during 2014) the older group were aged between 53 and 60, and the younger group between 18 and 32. Though there are only four older speakers (and one of them is often an outlier, as will be demonstrated below), this broad comparison is meant (a) as an indication of the norms of older SAfE, specifically regarding the short front vowels investigated here, and (b) to compare these to the norms of younger speakers. Given the discussion of the features of White South African English in Chapter 2 (section 2.3.1), it is evident by inspection of the data that the older speakers are representative of the norms described in the literature. In particular, there is evidence of retracted KIT (in all contexts), high DRESS and fairly high TRAP.

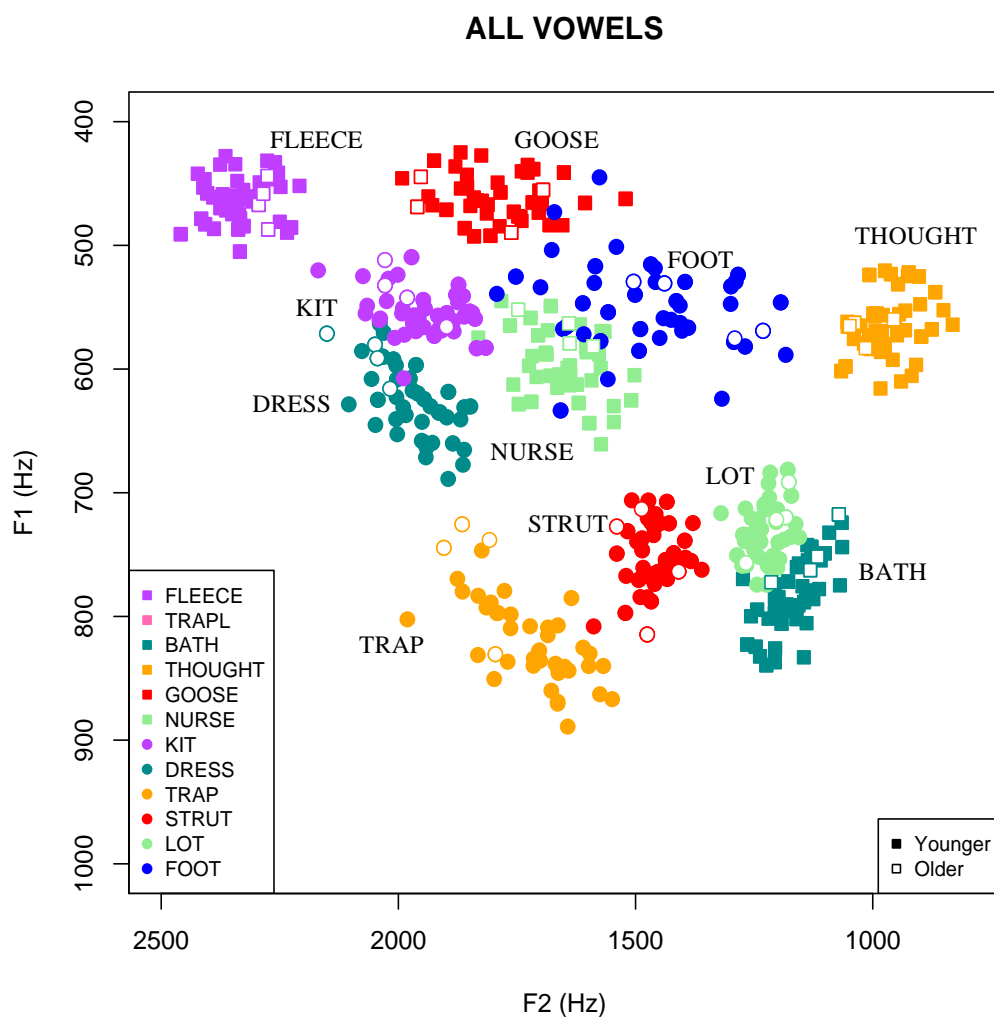


Figure 4.2: Means per speaker of each monophthong (in all environments) in the sample of Capetonian English speakers, separated by age.

The visual differences between the younger and the older speakers therefore indicate potential changes to the vowels of SAfE, which include the following for the long vowels:

- a) Fronter realisations of FLEECE.
- b) Raised and backer realisations of THOUGHT.
- c) Lower and fronter realisations of BATH.
- d) Lower realisations of NURSE.

The fronting of FLEECE is likely in response to fronted GOOSE to avoid merger. The raising and backing of THOUGHT is of particular interest, and is dealt with in Chapter 5. BATH has enjoyed recent attention, and in Cape Town, is found to be lower and fronter for women than for men, who fall into the category of speakers who display backing and raising (Mesthrie, Chevalier & Dunne 2015, 18). Conversely, the apparent time inferences drawn from figure

4.2 show that BATH is lower and fronter for younger speakers than it is for older speakers. Though these features are not the focus of this study they are certainly worth future attention.

Regarding the short vowels, figure 4.2 shows the following possibilities in terms of change in progress in the speech of younger speakers:

- a) Lowering and retraction of KIT.
- b) Lowering and retraction of DRESS.
- c) Lowering and retraction of TRAP (both long and short).

If these indications are viewed as an initial venture into the status of present-day SAfE, they suggest a the reversal of the Older Short Front Vowel Shift (henceforth Older Vowel Shift) in the lowering of KIT, DRESS and TRAP. These indications are merely impressionistic, and require statistical testing. Despite the uneven spread of the data regarding age, the statistical tests to follow include age as a continual variable in order to tease out differences in a statistical manner. The apparent time inferences drawn from the data would therefore need to be tested with a bigger generational sample. Furthermore, any fine-grained statistical results that emanate from the uneven sample are treated broadly when deemed not sociolinguistically relevant.

Figure 4.3 shows the mean values for all speakers for the short vowels, with standard deviation ellipses to indicate the spread of each vowel category. A mixed-effects model provides a numerical indication of the spacing of the short vowels in the vowel space. The model takes all tokens into account, and calculates how each vowel relates to TRAP as the intercept.⁵¹ The model considers height (F1) and frontness (F2) separately. In comparison to TRAP, all short vowels are significantly different in height and frontness ($p < 2e-16$). Along the F1 plane, Capetonian speakers have a 280 mean Hertz difference between their highest and lowest short vowels. KIT (545.6 mean Hz) and FOOT (543.6 mean Hz) are the high vowels; STRUT (760.6 mean Hz) and LOT (740.3 mean Hz) are the most mid vowels, and DRESS is a mid-high vowel (627.6 mean Hz). TRAP is the lowest vowel in the system at 825.6 mean Hertz.

⁵¹ TRAP is the intercept since it occurs first alphabetically as AE in the ARPAbet notation.

ALL SHORT VOWELS

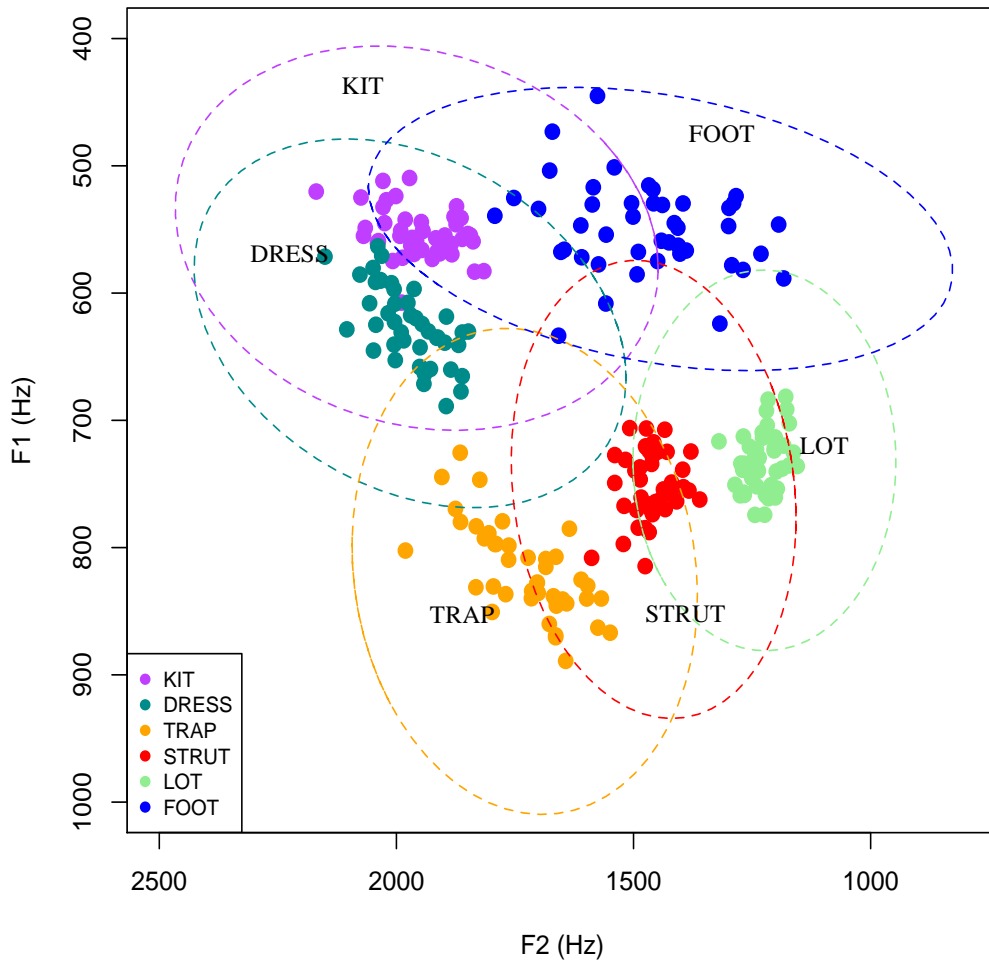


Figure 4.3: Mean values of all short vowels per speaker with standard deviation (x2) ellipses.

Along the F2 plane, Capetonians have a 782.8 Hz mean difference between the most advanced (DRESS) and most retracted (LOT) short vowels. DRESS is the frontest vowel (1973.4 mean Hz), followed closely by KIT (1921.9 mean Hz). TRAP is a front to central vowel (1733 mean Hz), followed by STRUT (1445.9 mean Hz) and FOOT (1384.7 mean Hz) as the central vowels. LOT is the backest short vowel in the system (1190.9 mean Hz).

Though overlap between categories is not abnormal (particularly since no outliers were excluded) the large overlap between KIT and DRESS warrants further analysis (provided in Chapter 5, section 5.3). TRAP and STRUT also overlap, although not as much as expected from merging vowels. In fact, STRUT overlaps to the same degree with LOT as it does with TRAP. Bekker (2009, 358-9), in addition to Lass (1990, 2002) noticed the large overlap between

FOOT and KIT occurring before /l/, which possibly explains the apparent encroaching of FOOT on the space occupied by KIT.

Figure 4.4 is a conditional inference tree, showing how the short vowels relate to one another according to height. As per node 1, the first significant ($p < 0.001$) binary split is between TRAP, STRUT, LOT (AE, AH, OH); and DRESS, KIT, FOOT (EH, IH, UH). Based on the positioning of the boxplots and their spread and medians, TRAP, STRUT and LOT are the ‘non-high’ vowels, while DRESS, KIT and FOOT are the ‘high’ vowels.⁵² The ‘high’ and ‘low’ groups are subsequently divided, each vowel significantly different from the other vowels.

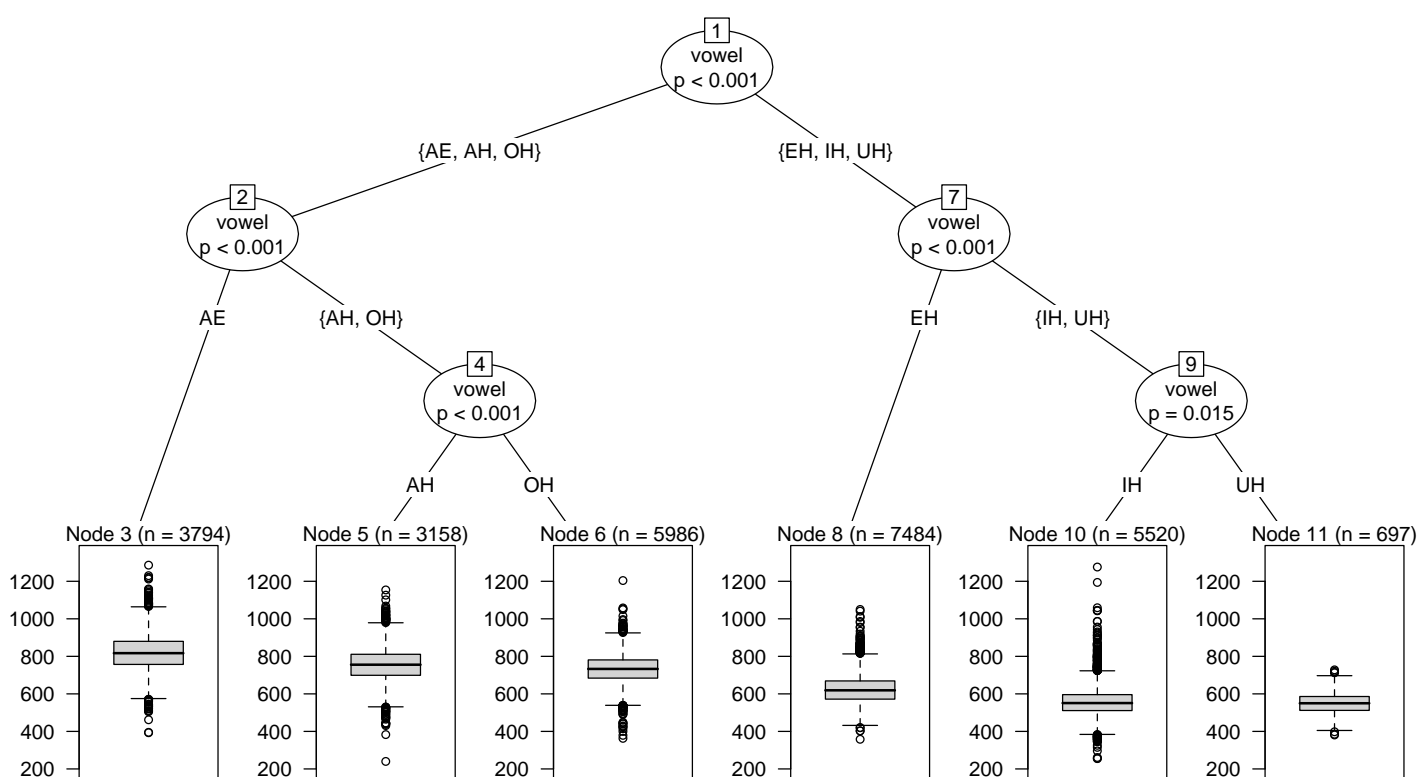


Figure 4.4: Conditional inference tree for F1 (height) of all short vowels.

The ctree also confirms what the mixed-effects model indicated: TRAP is the lowest vowel (node 3). This is evident in the median of over 800 Hz, as well as the spread of the upper quartile of tokens to over 1000 Hz. The highest vowels are KIT and FOOT, which centre under 600 Hz, though KIT is spread along F1 much more than FOOT. STRUT and LOT pattern together initially (node 2), as do KIT and FOOT (node 7) though ultimately not significantly so.

⁵² Evident by comparing the median line and the overall spread of the tokens in the box plots.

Figure 4.5 is the conditional inference tree for F2, which shows that all the vowels, as for F1, are statistically different from one another. TRAP, DRESS and KIT (AE, EH, IH) group together as the ‘front’ vowels, and STRUT, LOT and FOOT (AH, OH, UH) group together as the ‘back’ vowels ($p < 0.001$). STRUT and FOOT group together along F2 (node 8), though this is of little consequence since they do not group together for F1 (node 4, figure 4.4), negating the possibility of a merger. The lmer results are also confirmed: DRESS and KIT are the frontest vowels, centring around 2000 Hz along F2, though the spread of KIT (node 5) in comparison to the spread of DRESS (node 4) indicates that it is generally the frontest vowel. LOT is the backest short vowel with a median of about 1300 Hz (node 9). Though vowel categories overlap they are essentially statistically different from one another: the statistical significances indicate that mergers are not present (though these are broad observations).

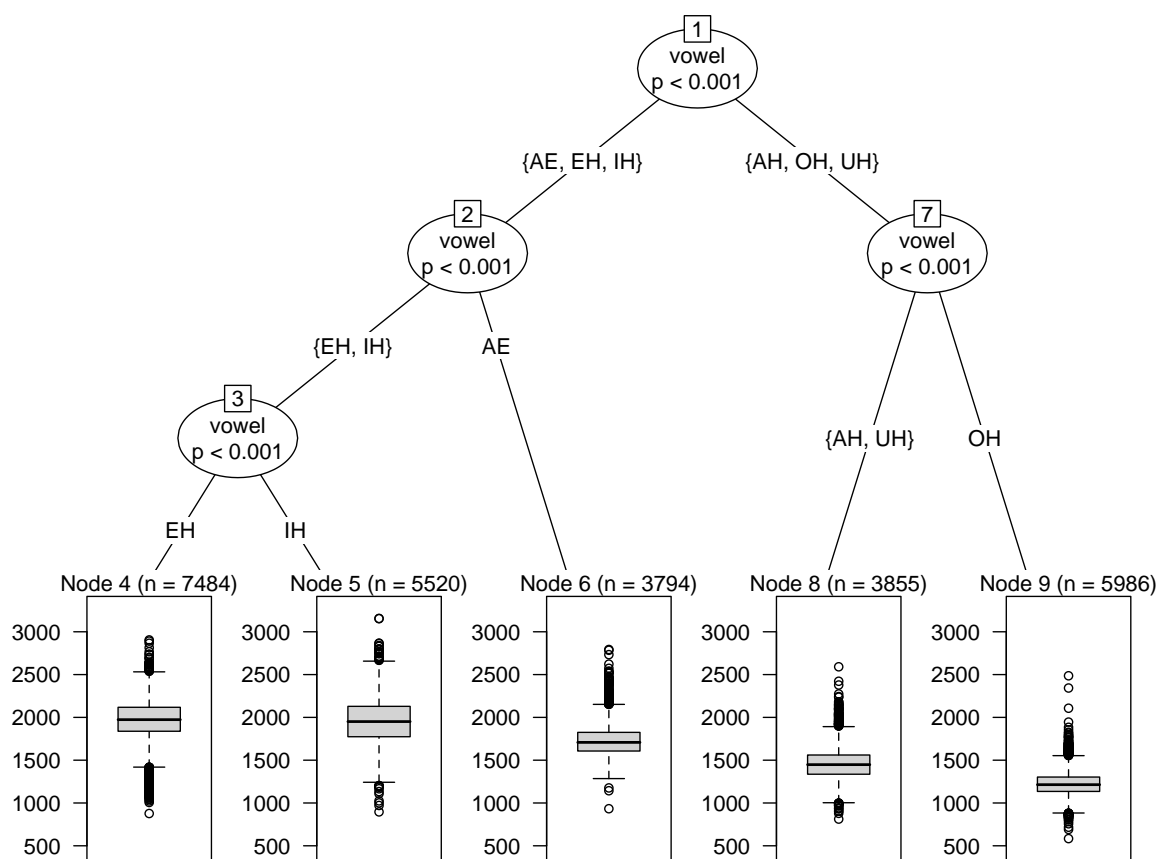


Figure 4.5. Conditional inference tree for F2 (frontness) of all short vowels.

The above descriptions have largely ignored the effects of the independent variables / fixed-effect predictors on vowel quality and variation, particularly phonetic environment and gender. The following sections thus explore the data further to tease out interactions between vowel quality as the variable; and ‘date of birth’, ‘gender’ and ‘environment’ as predictors.

4.2 THE KIT /ɪ/ VOWEL

All tokens of the KIT vowel were categorised based on the allophones identified by Bekker (2009, 2014a). Where he uses KIT4 to indicate KIT in the free position, I use KIT0 to mark it as the intercept for the mixed effect models.⁵³ I also do not consider the syllabicity of words (Bekker’s KIT1), choosing instead to rank primary vowel stress as more important than the number of syllables. In total, 5520 tokens of KIT were analysed after various other exclusions (discussed in Chapter 3, section 3.4.4), and per environment, are represented as follows:

Allophone	n	Environment	Example
KIT0	992	Free position	<i>fit</i>
KIT2	3412	Word initial	<i>it</i>
		After /h/	<i>hit</i>
		Before or after velars	<i>king</i>
KIT3	91	Before or after palato-alveolars	<i>stitch</i>
KIT5	832	After /l/ and /r/	<i>lit, rid</i>
		Before or after bilabials	<i>tip</i>
KIT6	193	After /w/	<i>wit</i>
		Before /l/	<i>till</i>
TOTAL:	5520		

Table 4.1: The allophones of KIT present in the data.

The KIT split was discussed in Chapter 2, i.e. the split of KIT into a fronter allophone before velars, word initially and after /h/ (KIT2); and a centralised allophone elsewhere. SAfE is the only English variety to treat KIT in this way. At first glance, figure 4.6 does not indicate a dichotomous split of the allophones: KIT2 (light blue) is indeed the most front allophone, but it overlaps considerably with the other environments. KIT6 is the most centralised, while the other environments occur in between. A linear mixed-effect model suggests that KIT0 occurs around 1900 mean Hz, and that all the other environments except for KIT2 are retracted in comparison. Of these, only KIT5 is significantly retracted ($p=0.000$). KIT2 is significantly fronter than KIT0 by 89 mean Hz ($p<0.000$). This is an indication of a KIT split, at least statistically, given that KIT2 is the most front realisation of the allophones. This supports Bekker’s (2014) notion of stabilised allophonic variation. The lmer shows no significant differences between the environments regarding height.

⁵³ ‘Free position’ as it is used here and henceforth, indicates the ‘elsewhere’ condition; i.e. when the vowel in question occurs in a phonetic environment that does not match the other identified environments.

KIT BY ENVIRONMENT

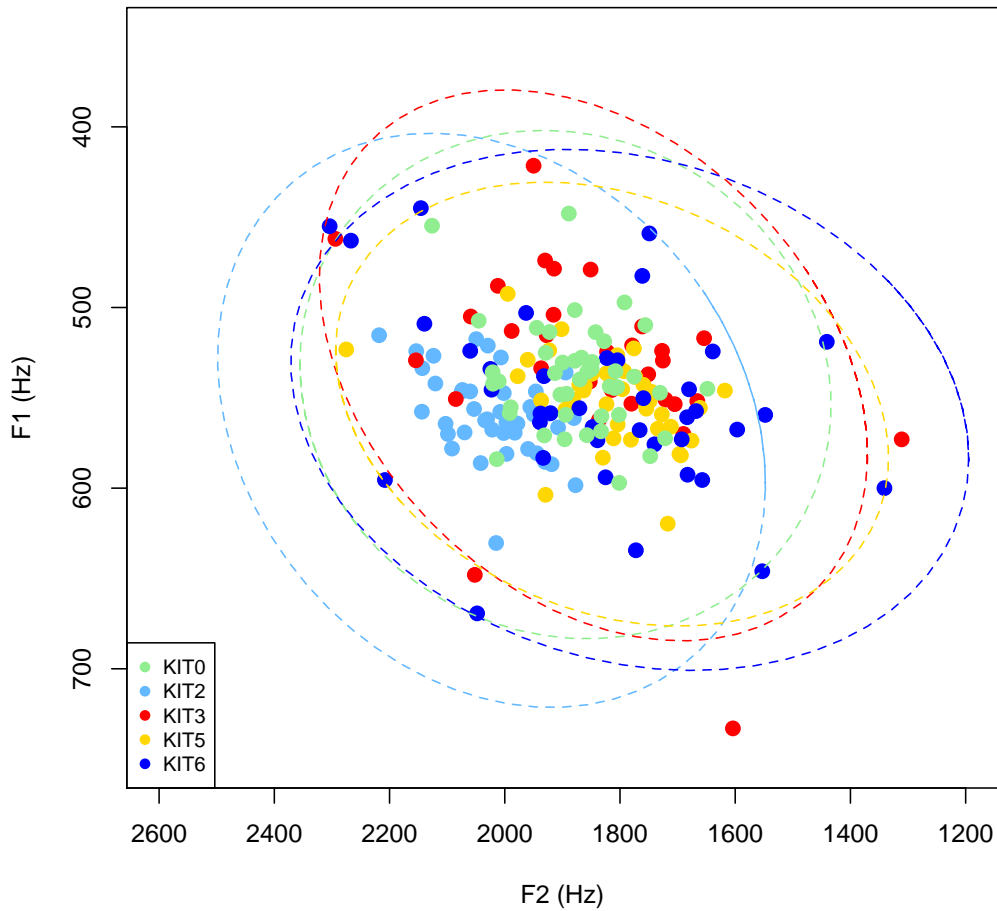


Figure 4.6: Mean values and standard deviation (x2) ellipses for the KIT allophones for all speakers.

The conditional inference tree for KIT F2 presented in figure 4.7 corroborates the results of the mixed-effects regression. Environment emerges as the most important predictor of frontness, significant at $p < 0.001$. The nodes encompassing KIT2 (3, 5 and 6) have medians that are fronter than the nodes encompassing the other environments (9, 10, 12 and 13), showing that KIT is allophonically and statistically split between front (KIT2) and retracted (elsewhere) realisations. Contrary to Bekker's (2014) claims, the KIT split, in the sense of a binary split between a fronter allophone and a set of more centralised allophones, is a feature of this variety of SAfE.

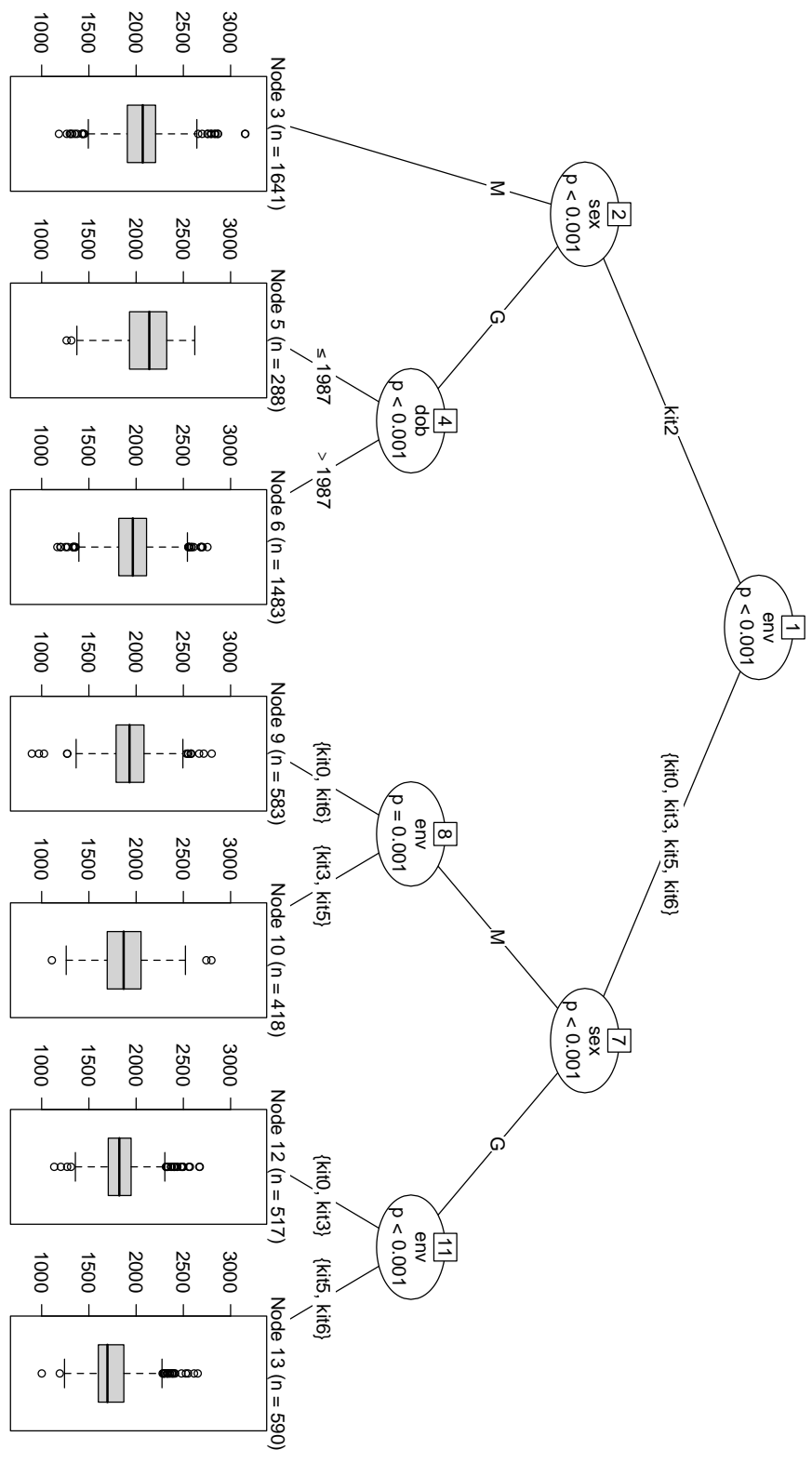


Figure 4.7: Conditional inference tree for KIT: F2 for 'environment', 'gender' and 'date of birth'.

After ‘environment’, ‘gender’ is an important predictor of F2. There is a significant difference in the way KIT2 is realised by men and women (node 2). The men are variable in their realisations of KIT2, evident in the tokens (‘outliers’) above and below the whiskers. Women show significant ($p < 0.001$) age differences: those over the age of 26 (node 5) have fronter (and historically more stable) realisations of KIT2 than women under 26 (node 6). This statistical divide between the older and the younger women provides some apparent time evidence suggesting that all young women (barring one exception) retract KIT2. Furthermore, the young women lead the change since all men generally have fronter realisations.⁵⁴ The speakers also lower the vowel, as I discuss shortly below.

The results for the other environments are not as clear, though ‘gender’ is also the most important predicting factor (node 7). Women treat KIT0 and KIT3 together as fronter than KIT5 and KIT6. Men, on the other hand, group KIT0 and KIT6 together, fronter than KIT3 and KIT5. KIT6 is expected to be the most retracted since it includes final /l/, before which KIT has been noted to retract to [ɤ]. The behaviour of the men is therefore an anomaly. It may be that since these environments undergo further retraction, final /l/ is as a result no longer a backing environment for all speakers. In all instances, however, women retract KIT ‘elsewhere’ more than men.

The conditional inference tree in figure 4.8 shows that ‘environment’ is also the most important predictor of F1, statistically significant at $p < 0.001$ (node 1). The environments are significantly grouped together: KIT2, KIT5 and KIT6 on the left, and KIT0 and KIT3 on the right. The height of KIT in the environments KIT2, KIT5 and KIT6 is significantly predicted by ‘date of birth’ (node 2). Regarding KIT2 and KIT6, older speakers have a higher median realisation than younger speakers (nodes 5 and 7 respectively).⁵⁵ This, together with the high degree of variation in node 7, suggests that KIT2 and KIT6 show lowering in apparent time.

⁵⁴ MarilynK, though a young woman born in 1987, patterns with the 3 older women in the sample despite a 26-year age gap. In this way, MarilynK maintains a more traditional realisation of this vowel.

⁵⁵ This excludes KayC, an older woman born in 1961, who patterns with the younger speakers. She was a first year student at UCT, and therefore has contact with younger speakers to which she likely accommodates.

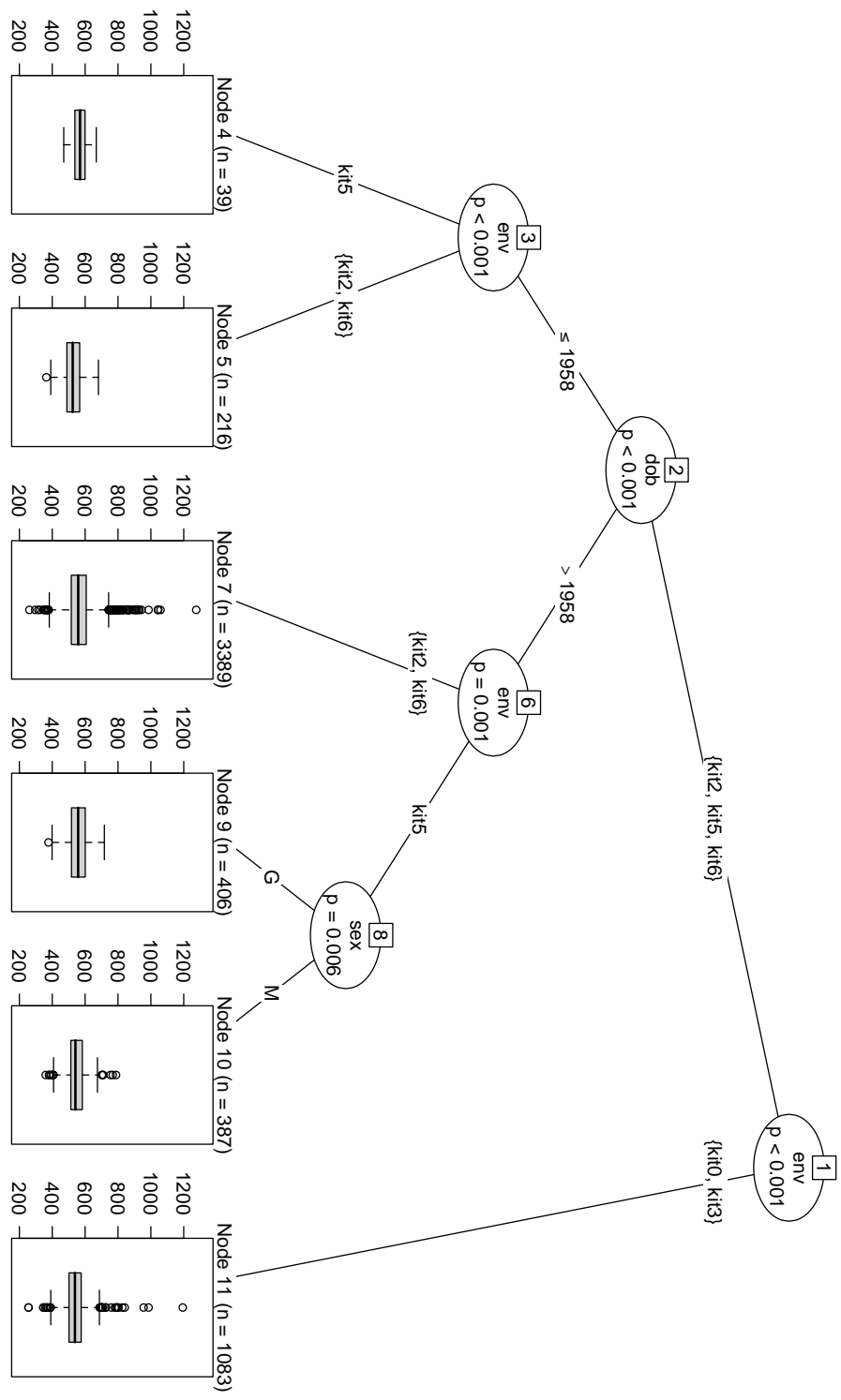


Figure 4.8: Conditional inference tree for KIT: F1 for 'environment', 'gender' and 'date of birth'.

KIT5 is stable for the older speakers (node 4), though for younger speakers there is a gender difference ($p=0.006$): women produce KIT5 lower than men do (node 9 and 10 respectively). This could be the start of lowering KIT in a new environment (i.e. in addition to KIT2 and KIT6), particularly in the speech of young women.⁵⁶

KIT0 and KIT3 group together (node 11) as significantly different from the other environments, though not different to one another, and not predicted by speaker gender or age. The number of tokens occurring above the whiskers indicates a lowering trend by all speakers in this environment, though these allophones are generally higher than the others.

In SAfE spoken in Cape Town, therefore, KIT is undergoing backing and lowering, the details of which are summarised in table 4.2. Though a kind of KIT split is present, young women retract KIT2, which is the environment where KIT traditionally remains more front. Despite this retraction, the KIT split is maintained since the remaining allophones are undergoing secondary retraction, i.e. further retraction of the already centralised allophones.

Allophone	Change	Speakers
KIT0	Lowering Retracting	All, incipient Particularly women
KIT2	Lowered Retracted	All young Young Women
KIT3	Lowering Retracting	All, incipient Particularly women
KIT5	Lowering Retracting	All young women Particularly women
KIT6	Lowering Retracting	All young speakers Particularly women

Table 4.2: Summary of changes to KIT in SAfE.

The overall tendency towards the lowering of KIT provides the first shard of evidence for the Reverse Vowel Shift: new KIT is becoming lower than old KIT, a change led by young women. This provides evidence in support of Mesthrie's (2012a) claims for the lowering of KIT. In addition to this lowering, retraction of KIT and its two allophones (fronter KIT2 and backer 'other' KIT) is evident.

⁵⁶ Again, KayC included in the group of 'lowerers', who are at least 26 years younger than her. She is accommodating her speech to the norms she is currently exposed to as a first year student at university, where she is surrounded by people between the age of roughly 19 and 21.

Lanham (1965) predicted, as part of the Older Vowel Shift, the continued centralisation of KIT. This points to the possibility that the younger speakers in my study are advancing this aspect of the Older Vowel Shift. In a strict sense, it is therefore terminologically problematic to call particularly the backing of ‘other’ KIT part of a reversal of the Older Shift (especially since the centralisation of ‘other’ KIT occurs instead of fronting as would be expected of a full reversal of the Older Shift). I however maintain that the changes to KIT are part of the ‘Reverse Vowel Shift’ despite the terminological mismatch. In the lowering of KIT2 AND ‘other’ KIT, the Older trend of raised KIT is undone. The further centralisation of ‘other’ KIT results in the maintenance of the KIT split, a distinctly South African feature. Though the centralisation of ‘other’ KIT runs counter to the notion of a Reverse Vowel Shift, it occurs as a result of the Reverse Vowel Shift, and it is therefore part and parcel of it.

4.3 THE DRESS /e/ VOWEL

In recent socio-acoustic studies of varieties of SAfE (Toefy 2014, Bekker 2009), the DRESS vowel was analysed in terms of two environmental positions: (1) DRESS before /l/, where it is known to lower to /æ/ with retraction in especially White South African English (Chapter 2, section 2.3.1); and (2) DRESS occurring elsewhere. In addition to these environments, Thomas (2001) and Labov (1994) outline general conditioning environments for short vowels. If DRESS in SAfE behaves as short vowels in Englishes tend to, we can expect the following phonological interactions:

1. DRESS followed by velars is lower and fronter than DRESS in the free position (Thomas 2011, 101).
2. DRESS followed by nasals is fronter than DRESS in the free position (Labov 1994, 275).
3. DRESS retracts before final /l/ (Thomas 2011, 126).

The DRESS tokens were therefore coded for phonetic environment accordingly, shown in table 4.3. Manner of articulation for preceding and following segments are not considered here in order to facilitate comparisons to earlier studies, though it merits future investigation. After various exclusions, 7484 tokens were available for analysis.

Environment	n	Example
_0	4875	<i>bet</i>
_g/k	614	<i>egg, peck</i>
_l	312	<i>bell</i>
_m/n/ŋ	1683	<i>sense, pen, length</i>
TOTAL:	7484	

Table 4.3: Number of DRESS tokens per environment.

As per the descriptions of DRESS in Chapter 2, it is a raised, somewhat centralised vowel, and the general realisation is in the region of [e] or slightly more centralised [ë], as per the Older Vowel Shift. The vowel is also, as mentioned earlier, characterised by extreme lowering and backing to [æ] before /l/.

Figure 4.9 is the formant plot for DRESS in the various environments. There is considerable overlap, though backing and lowering of the ellipses is evident. DRESS_0 (light blue) is the most front variant, and DRESS retracts and lowers based on the phonetic nature of the following segment. At face value, apart from retraction before final /l/, the formant plot suggests that Capetonians do not treat DRESS in the same way phonologically as speakers of other Englishes. Where we expect nasals and velars to cause fronting and lowering, the formant plot in figure 4.9 shows lowering and retraction.

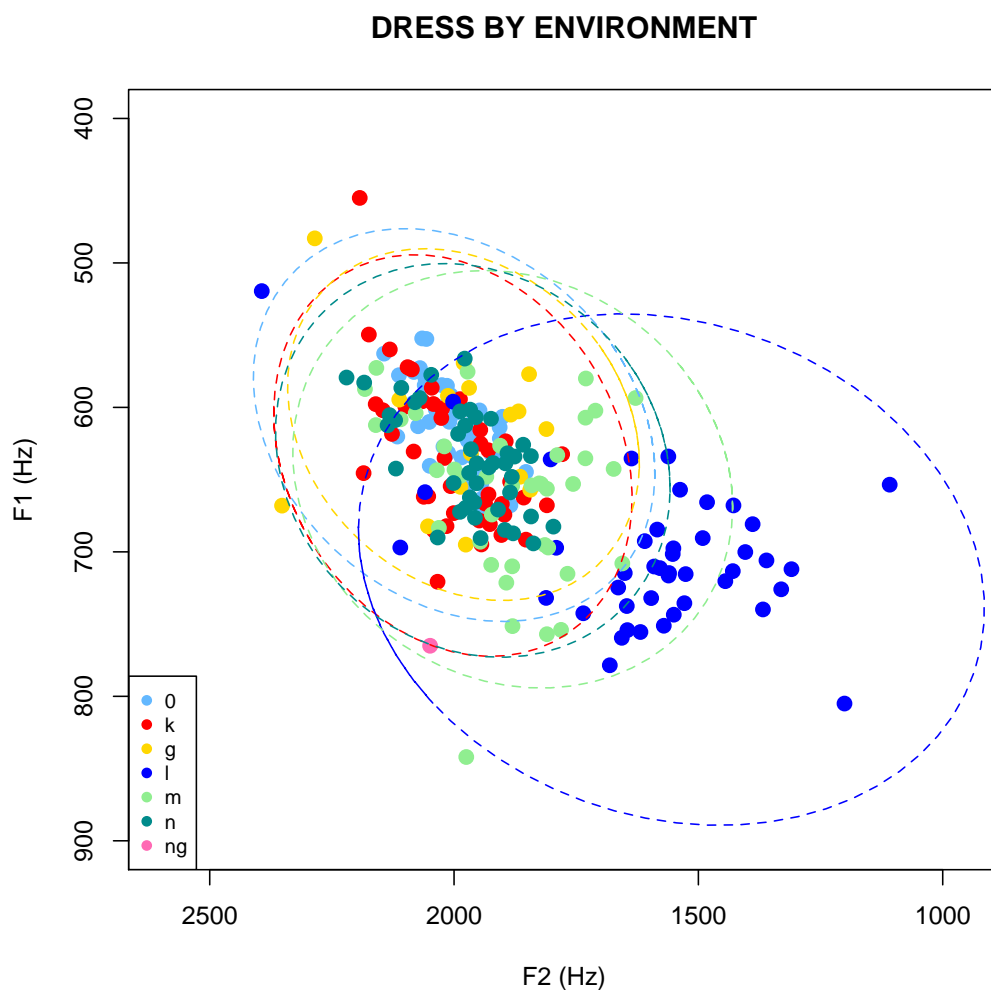


Figure 4.9: Mean values and standard deviation (x2) ellipses for DRESS for all speakers by environment.

Vowel	Estimate	SE	<i>t</i> -value	<i>p</i> -value
F1				
0 (int)	620.33	4.710	131.72	<2e-16
g	0.60	12.469	0.05	0.962
k	12.57	4.081	3.081	0.002
l	107.74	6.67	16.15	<2e-16
m	37.18	8.55	4.35	1.51e-05
n	11.21	3.47	3.23	0.001
ng	102.76	62.47	1.65	0.100
F2				
0 (int)	1983.86	11.23	176.63	<2e-16
g	-16.49	41.03	-0.40	0.688
k	11.90	13.39	0.89	0.375
l	-459.35	22.06	-20.83	<2e-16
m	-10.11	28	-0.36	0.718
n	27.8	11.45	2.428	0.015
ng	102.63	198.16	0.52	0.605

Table 4.4: Mixed-effects regression results for DRESS.

The results of a mixed-effects model (table 4.4) clarify the interactions between DRESS_0 and the other environments. In relation to the intercept (DRESS_0), velar and nasal environments cause DRESS to lower. However, lowering is only significant for DRESS_k, DRESS_m and DRESS_n. Fronting is also only significant for DRESS_n: velars and the other nasals do not cause significant fronting. DRESS_l is lowered and retracted to such an extent that it is an allophone significantly different to the others.

The way in which the phonetic environments relate to one another, and to the social factors gender and age, are better ascertained via conditional inference trees. Too big to reproduce, these trees selected ‘environment’ as the most important predictor of both F1 and F2, and DRESS_l branches off as a separate allophone ($p < 0.001$ for F1 and F2). In contrast to the lmer results, a ctree suggests that final /l/ has no significant influence on the height of DRESS. Retraction of DRESS before final /l/ is predicted by ‘gender’ and age, as seen in figure 4.10. Women born after 1987 (node 5) have significantly ($p = 0.013$) retracted realisations compared to women born in or before 1987 (node 4).⁵⁷ Men show no internal differentiation. Based on these results, DRESS_l is undergoing secondary backing, led by the young women; ‘secondary’ in the sense that DRESS_l is already a retracted vowel, retracted further by these

⁵⁷ Again, MarilynK patterns with the older women, and does not participate in secondary backing along with her peers, maintaining an older norm.

young women. Since DRESS_1 behaves in a way quite different to the other environments, it is appropriate to exclude it from the analysis for the time being.

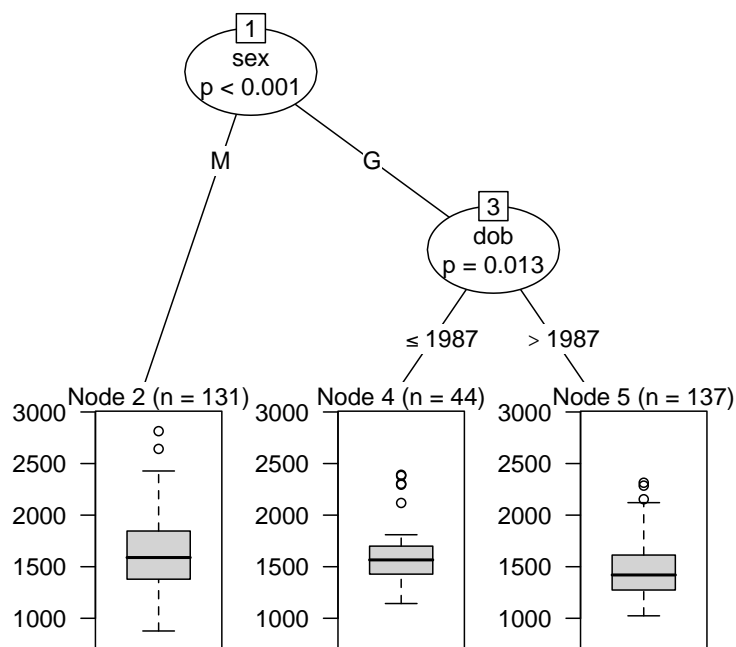


Figure 4.10: Conditional inference tree for DRESS before /l/: F2 for ‘gender’ and ‘date of birth’.

After the exclusion of DRESS_1, a conditional inference tree shows that ‘gender’ is the most important predictor of F1 ($p=0.001$). Figure 4.11 shows the conditional inference tree for DRESS for just the men. Men over the age of 26 (node 2) draw a distinction between DRESS before nasals and DRESS in the free position / before velar plosives. That the velar and free environments group together is in contrast to the expectation (and the lmer results) that DRESS before velars is lower than DRESS in the free position. Instead, DRESS realised before nasals is lower than DRESS in the free and velar position. This can be gauged by a visual comparison of the medians of the box plots in nodes 4, 5 and 6, though mixed-effects modelling supports this (see table 4.4). Moreover, when DRESS is followed by /m/ it is lower than when it is followed by /n/ (nodes 4 and 5).⁵⁸

⁵⁸ The small number of DRESS_m tokens may have an effect on the results.

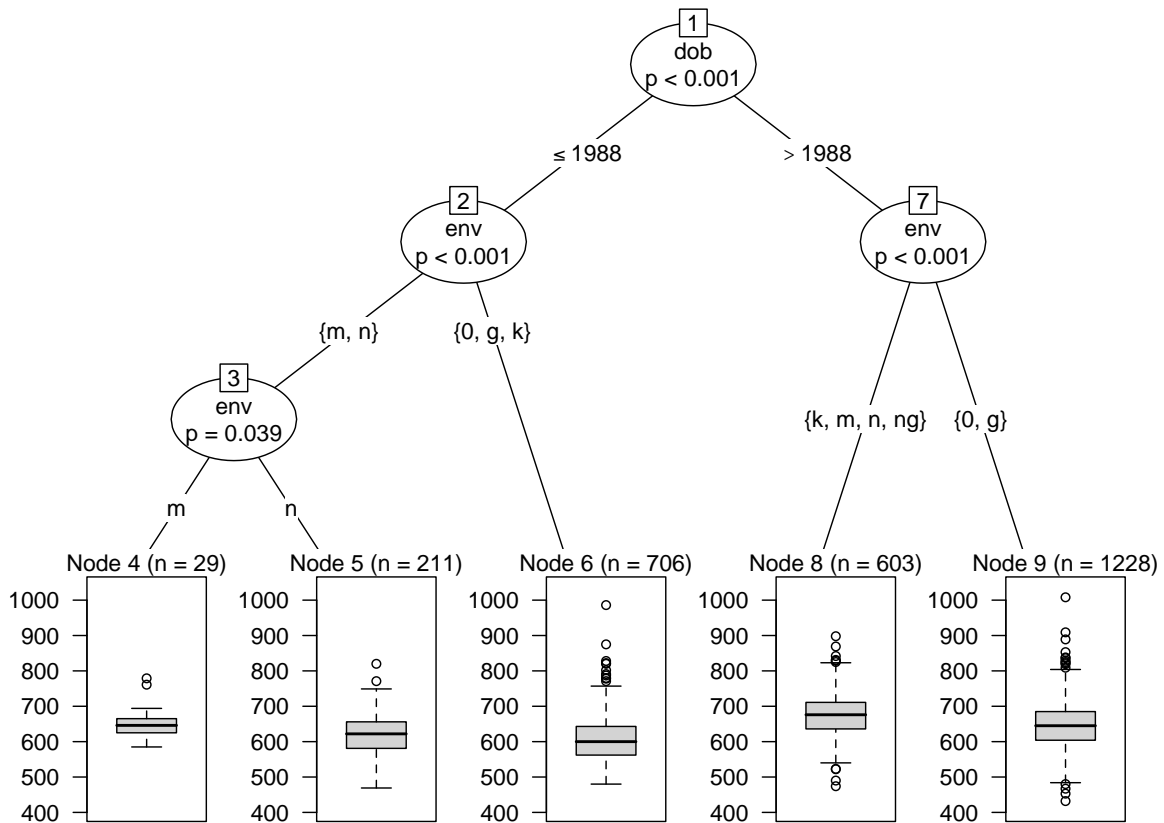


Figure 4.11: Conditional inference tree for DRESS (men): F1 for ‘environment’ and ‘date of birth’.

Men under the age of 26 (node 7) have altered the allophonic grouping of DRESS when compared to older men. DRESS_k is re-categorised as separate from DRESS₀ and DRESS_g, treated as similar to the nasals instead. When DRESS occurs in the free position or before /g/, it is higher than when it occurs before /k/ or nasals. After this re-categorisation, the men under 26 produce lower realisations in the ‘nasal’ category than men over 26 (compare nodes 4 and 5 to node 8). In addition, DRESS in the free and ‘velar’ category is also lower for the younger men (compare node 6 to node 9). In general, therefore, men under the age of 26 have lower DRESS realisations in all environments compared to men older than 26. This is a rather fine-grained statistical difference, especially given the fact that there is one truly older male in the group of ‘over 26s’, while the other members are at most five years older, and are therefore in the same generation as the ‘under 26s’. The results therefore suggest that DRESS, particularly before /k/ and nasals, but also in the free environment and before /g/, is being lowered by the young men in the sample. Phonologically we expect lowering before velars, thus the lowering before nasals and in the free environment implies reasons other than phonological.

The women, presented in figure 4.12, have a very similar relationship between DRESS and its environments, though ‘date of birth’ is a more statistically prevalent indicator. Women over the age of 24 (node 2) treat DRESS and its environments in the same way as the older men discussed above: DRESS in the free position is generally higher than when it is followed by nasals.

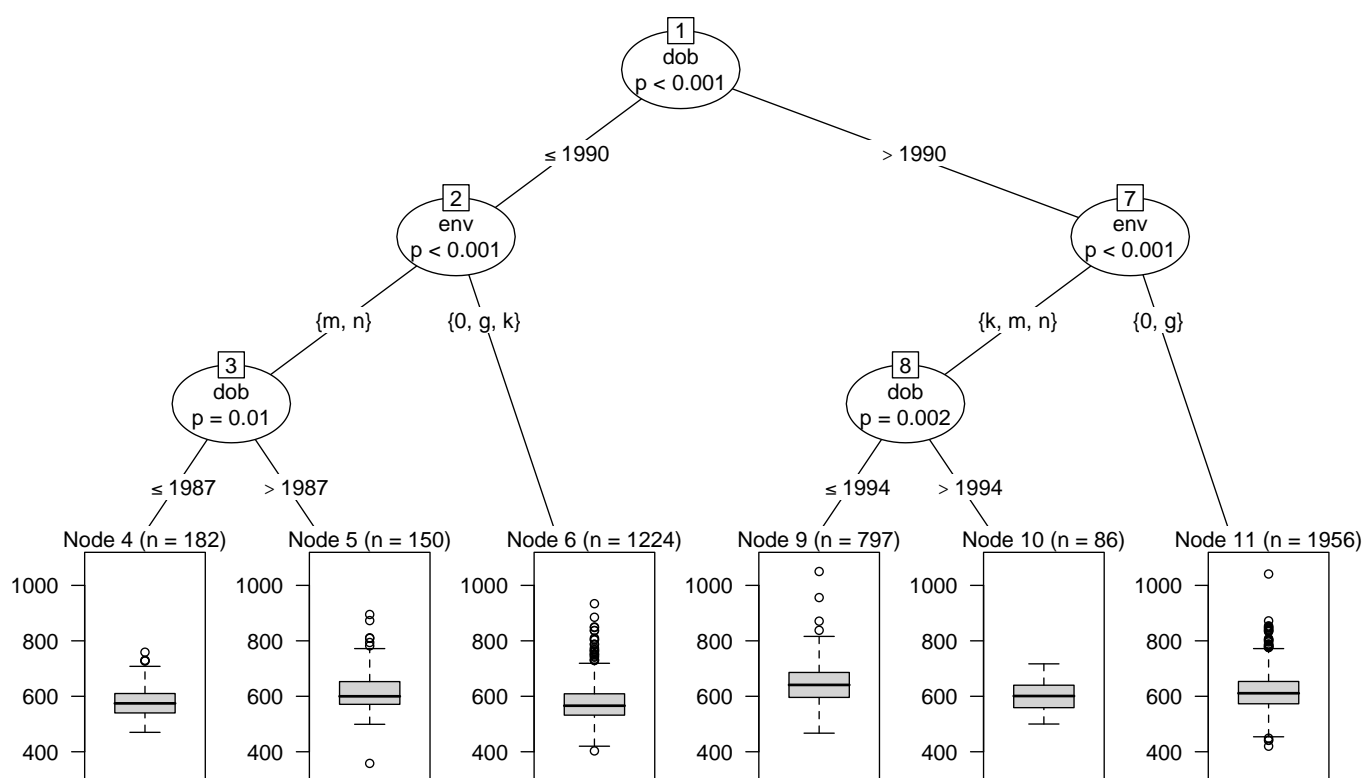


Figure 4.12: Conditional inference tree for DRESS (women): F1 for ‘environment’ and ‘date of birth’.

Women aged between 20 and 24 also re-classify DRESS_k, treating it as they treat nasals, realising it lower than the older women (compare node 3 and node 9).⁵⁹ In addition, DRESS in the free position or before /g/ is lower for women younger than 23 than it is for women older than 23 (node 6 vs. 11). Again, these statistical results are too fine-grained by cut-off years to indicate concrete significance for age, with members of the same generation appearing different. Broadly speaking, therefore, the DRESS vowel appears to lower in apparent time: the younger the speaker, the lower the vowel.

⁵⁹ With one exception: CindyH (1995, node 10) behaves differently to speakers her age and maintains an older, higher variant.

Though both genders lower the vowel, men have lower realisations than women (around 700 mean Hz for men compared to around 600 mean Hz for women along F1). That men lead change is not very common in sociolinguistics. Lass (2002, 115) however notes that women in general have higher DRESS realisations than men. It is therefore likely that a raised DRESS vowel remains a competing prestige variant appealing to women. DRESS therefore remains a gendered variant despite lowering.

The most important conditioning factor for the F2 of DRESS is ‘date of birth’ ($p < 0.001$). Speakers born in or before 1989 are differentiated from those born after 1989. Figure 4.13 shows the conditional inference tree for speakers 25 or older (born in/before 1989).

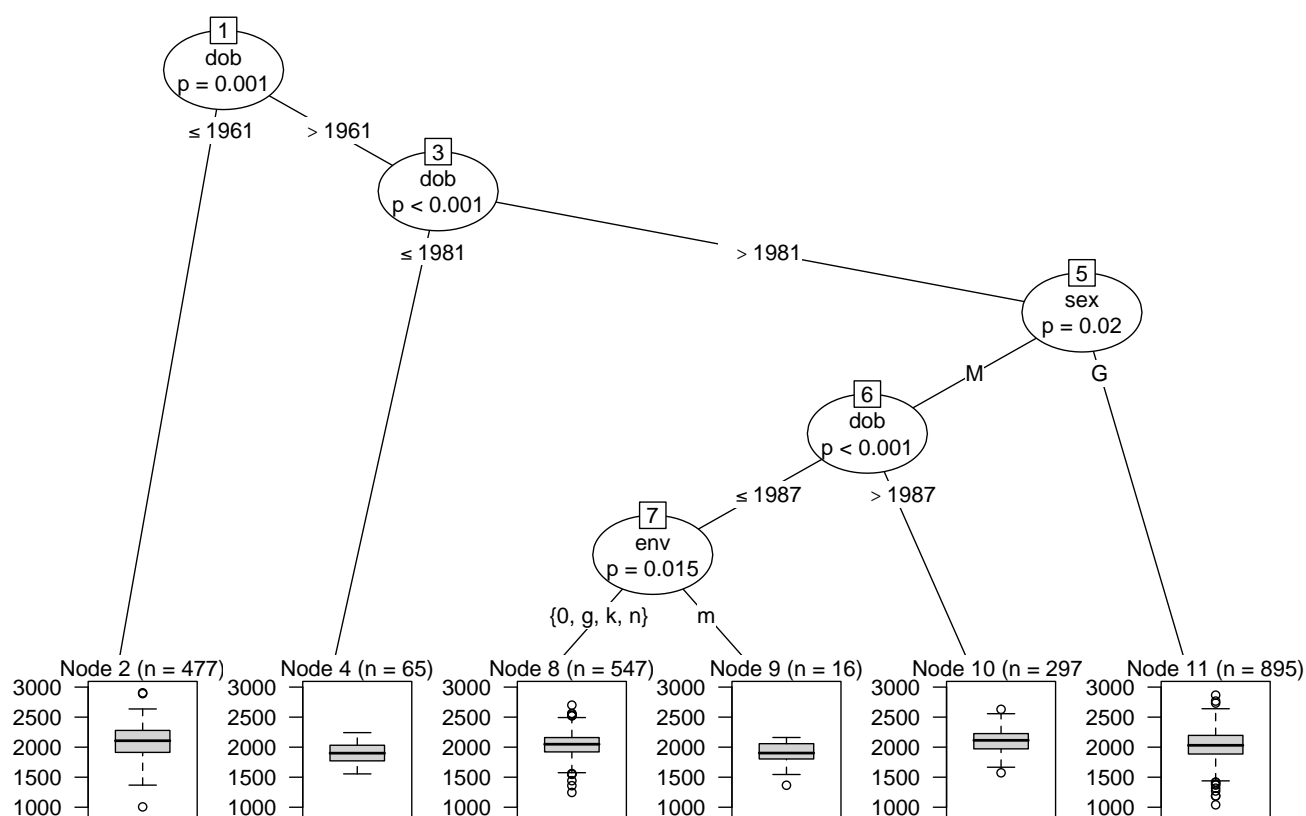


Figure 4.13: Conditional inference tree for DRESS: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1989.

The older speakers are grouped together (node 2), and generally realise DRESS as a front vowel, with a median well over 2000 Hz. For the younger speakers, DRESS is conditioned by ‘gender’ (node 5). Women do not show differentiation in F2 for DRESS, and generally it is also a front vowel with a median just over 2000 Hz. Men born after 1987 are more stable

along the F2 plane, and show a slightly fronter realisation of DRESS than women (compare node 10 and 11).

There is allophonic differentiation in the realisations of men born between 1983 and 1987 (node 7): DRESS_m is significantly retracted (node 9). This is also the environment that undergoes the most lowering for men of a similar age (node 4, figure 4.11). Only one speaker is represented by node 3: JimH, born in 1981. He is the only exceptional speaker displaying significantly retracted DRESS (below 2000 mean Hertz): only DRESS_m is retracted and lowered by the other men born before 1987.

The retraction of DRESS is more evident in speakers between 19 and 24 years of age (i.e. born after 1989). Figure 4.14 shows that in this age group, ‘environment’ is the strongest predictor of the F2 of DRESS. Where the median F2 of DRESS is generally above 2000 Hz for the speakers in figure 4.13, this group consistently shows a median F2 value of well under 2000 Hz. Nasals (barring /ŋ/, treated as velar) cause the most backing of the DRESS vowel, and are thus grouped together as separate from the other environments. Retraction is strongest when DRESS is followed by /m/ (node 7), as it is for men in the ‘older’ group. DRESS before /n/ is conditioned by gender, where women retract more than men (node 4). In short, the younger speakers in the sample, especially women, retract the DRESS vowel. Centralised variants of DRESS (i.e. [ɛ̈]) occur in the description offered by Lass (1990, 276). Given the further retraction present in the speech of young Capetonians, it would appear that this centralisation had stabilised, and is now being retracted further.

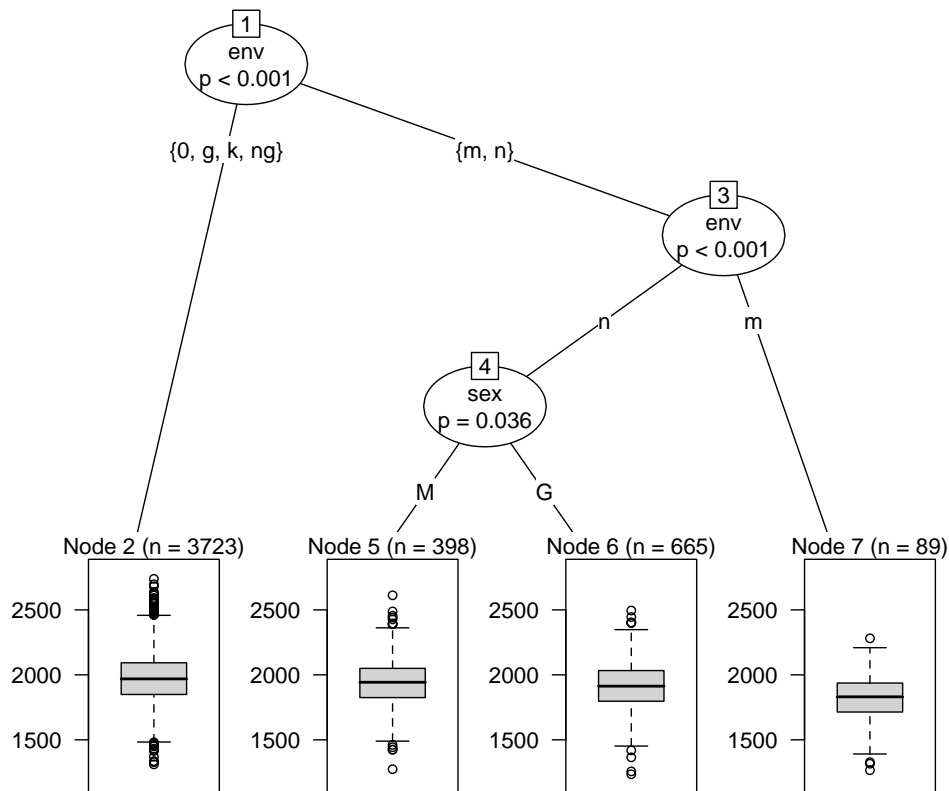


Figure 4.14: Conditional inference tree for DRESS: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1989.

DRESS in Cape Town’s variety of SAfE is lowering and retracting, the particulars of which are summarised as follows:

- a) All speakers lower DRESS_m and DRESS_n, though DRESS_m is lower (suggesting that this is the environment in which lowering started). Younger speakers extend this lowering to include DRESS_k and DRESS_ŋ. Generally, the younger the speaker, the lower DRESS is, and in more contexts than what is the case for the older speakers.
- b) Importantly, men have lower DRESS realisations than women. This can be seen as an attempt to maintain the raised variant perceived as prestigious by women (though the retraction of DRESS potentially moves the vowel away from this ‘prestigious’ quality).
- c) Overall, women lead retraction, with backer DRESS realisations than men.
- d) All younger speakers retract DRESS before nasals more than older speakers, particularly DRESS_m. In this sense it appears that this environment favours retraction (and lowering).

Thus far, the Reverse Vowel Shift is evidenced in the lowering of both KIT and DRESS, though secondary retraction before /l/ is an innovation. The centralisation of DRESS is a feature previously associated with SAfE, and can be linked to generalised sound change given prevalent KIT centralisation in this variety (see Labov 2010, 112).

4.4 THE TRAP /æ/ VOWEL

According to Thomas (2011) and Labov (1994), as outlined in section 4.3, vowels tend to be affected by certain phonological conditions. We therefore expect TRAP to retract before final /l/, front before nasals and lower and front before velars. In his study, Bekker (2009) included unmarked TRAP (*back, bag*), TRAP occurring before non-sonorant coronals (*bad, bat*), and TRAP before nasals (*ban, bang*). The data however shows “no evidence of allophonic variation” with regards these distinctions (Bekker 2009, 201). He does however show TRAP retraction before /l/. Bearing these environmental effects in mind, the distribution of the data per environment is therefore as follows:

Environment	n	Example
_0	2675	<i>bat, bad</i>
_k/g	1354	<i>back, bag</i>
_l	26	<i>alcohol</i>
_m/n/ŋ	1355	<i>camp, cancer, bank jam, tan, angry</i>
TOTAL:	5413	

Table 4.5: Number of tokens of TRAP by environment.

As a result of the Older Vowel Shift, TRAP in SAfE has been described as ranging between [æ], [ɛ] and [ɛ̄], depending on the speaker. Bekker and Eley (2007) and Bekker (2009) found an early indicator of the Reverse Vowel Shift in the retraction and lowering of TRAP in their data.

Figure 4.15 presents the formant plot of TRAP in its environments. Though the spread of TRAP_l looks rather different, lmer tests show that there is no significant lowering ($p=0.794$) or retraction ($p=0.950$) before final /l/. TRAP lowers significantly when it occurs before /k/ ($p=0.006$). It also lowers before /g/, though only approaching significance ($p=0.522$), perhaps indicative of the beginnings of a change. Before nasals /n/ and /m/, TRAP is fronter than in the free position, though not significantly so ($p=0.096$ and $p=0.168$ respectively). Thus far, the speakers condition TRAP largely in ways that are expected phonetically, except that final /l/ is not a significant backing environment and velars do not cause fronting. Furthermore, the mixed-effects results show that the tense (before voiced codas) and lax TRAP allophones are not statistically different from one another regarding quality.

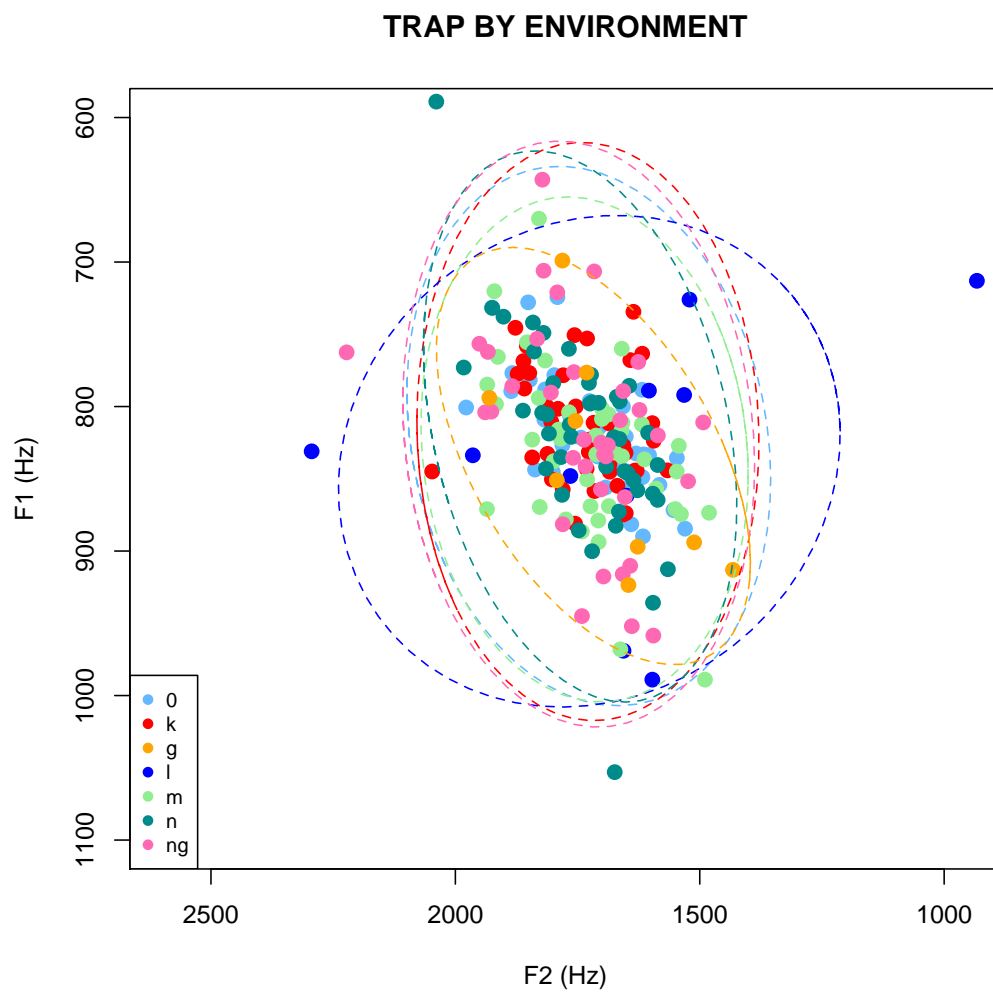


Figure 4.15: Mean values and standard deviation (x2) ellipses of the TRAP vowel and its environments.

A conditional inference tree selects ‘date of birth’ as the strongest predictor of TRAP for both F1 and F2. Speakers born between 1954 and 1987 are separated from those born between 1988 and 1995. Figure 4.16 presents the tree for the older group.

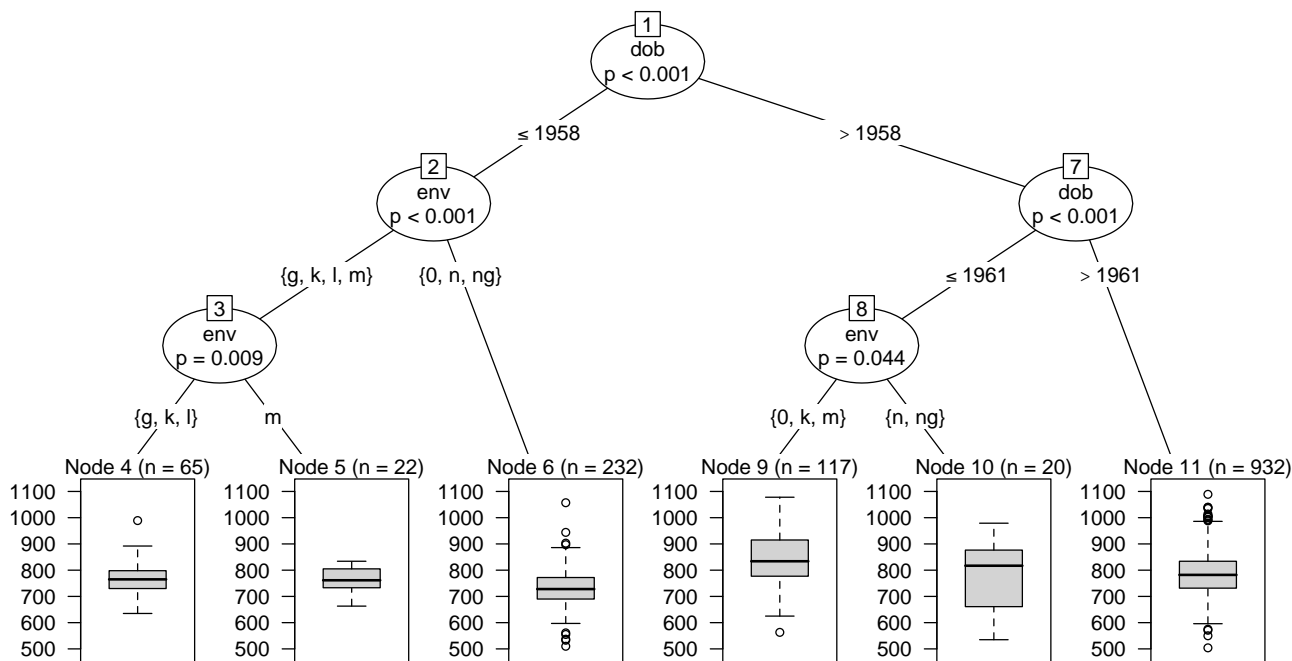


Figure 4.16: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1987.

Speakers between the age of 56 and 60 (node 2) realise TRAP variably based on phonological environment, evident in the medians of nodes 4, 5 and 6. The lowest realisations (node 4) occur before velars, a lowering environment. Compared to TRAP in the free position (node 6), TRAP_m is lower. Since nasals are not generally conducive to lowering, lowering in this environment can be seen as the start of lowering in all the environments. This is supported by the fact that speakers aged between 27 and 52 (node 11) show no environmental distinctions.

Figure 4.17 shows the results for speakers born between 1988 and 1995 (i.e. between 19 and 26 years of age). TRAP is clearly lowered for these speakers in all environments: their medians are above 800 Hz as opposed to the speakers in the older age group who centre just below 800 Hz.

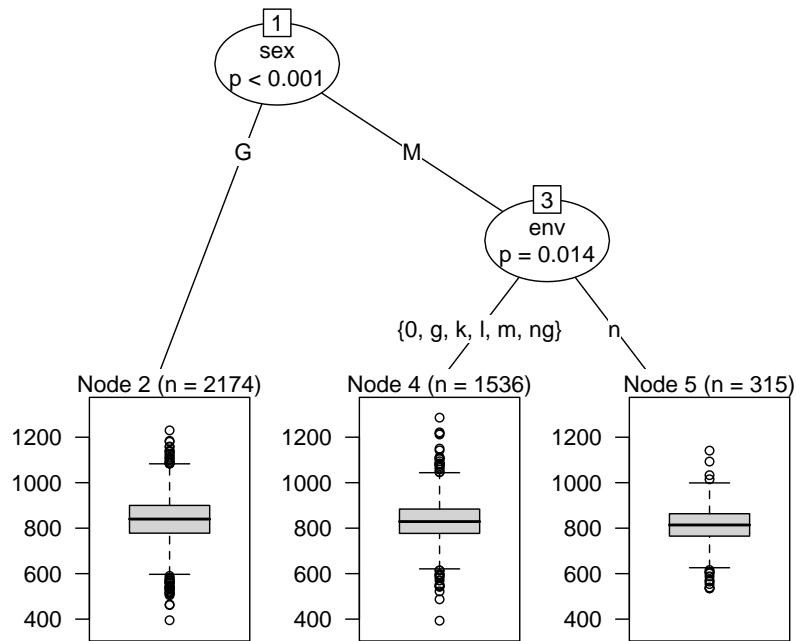


Figure 4.17: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1987.

While there are environmental distinctions in this age group, they are not the same as those observed for the older group: ‘gender’ is a more significant predictor. The women in this age group have the lowest realisations of TRAP (node 2). Men distinguish between a higher TRAP_n and a slightly lower TRAP realised elsewhere. Given that only one environment is lower, it follows that this environment lags in respect of the lowering experienced in the other environments, particularly given that women show no such differentiation. Comparing these results to those of the older set of speakers, it is evident that TRAP in all the environments is lowering, and that the lowering is in full effect for the younger speakers, both men and women.

Outlier

Node 8 in figure 4.16 represents KayC, who is placed in the 27 to 53 year old group despite being the only speaker to show environmental differences in her realisations of TRAP. Her realisations are also significantly lower than all the other speakers represented in this ctree. She is in this ‘in-between’ stage because she accommodates to the norms of the younger speakers that she is exposed to at university, who are represented in figure 4.17. At present she accommodates their height norms, but maintains the allophonic distinctions her

contemporaries make: TRAP is higher in the free position and before /ŋ/. In this sense, she accommodates to younger norms, though not fully.

The degree of frontness of TRAP is significantly predicted by ‘environment’ for speakers between 27 and 60 years of age (node 1, figure 4.18).⁶⁰ All environments, except TRAP_k,l are treated in the same way by all speakers, and is realised at a median of around 1900 Hz. When TRAP occurs before /k/ or /l/, one speaker retracts it to such an extent so as to cause a gender difference (node 7). Given that this is the only instance, and that TRAP in the other environments does not differ greatly from one another (compare nodes 2, 4 and 6) it follows that TRAP, overall, is largely stable for this age group.

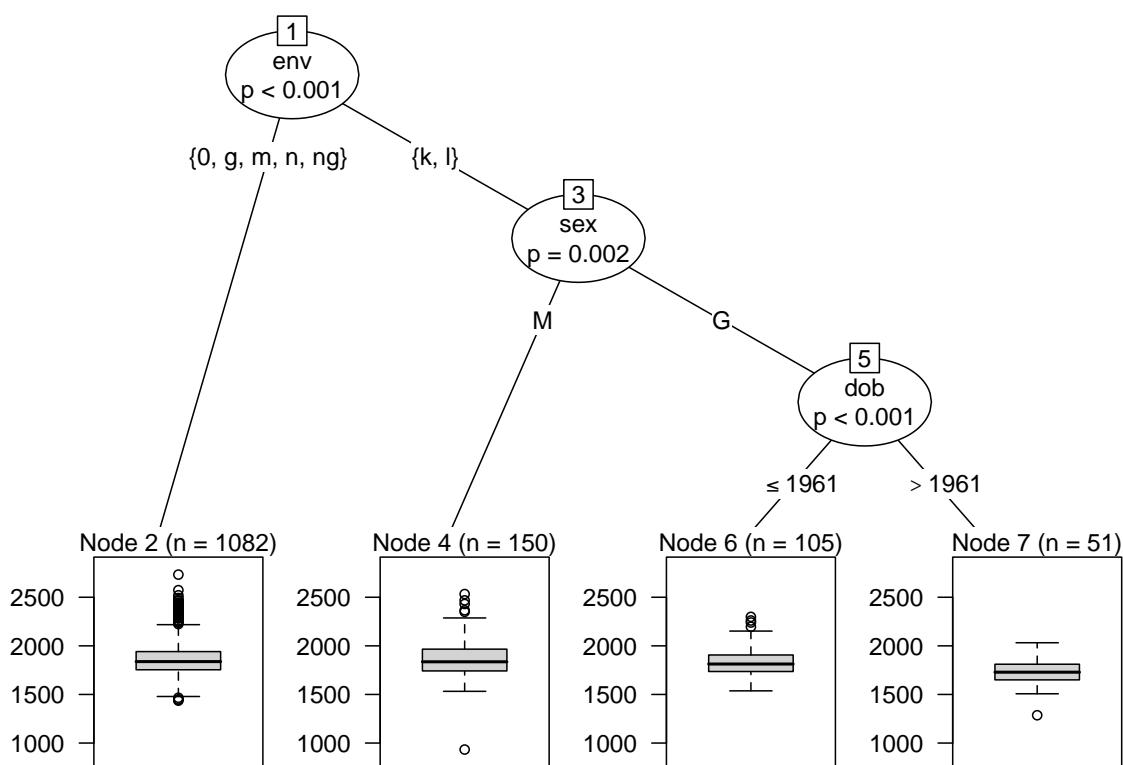


Figure 4.18: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born in or before 1987.

⁶⁰ MarilynK is again grouped with the older ladies instead of with her peers. The reason for her continuous patterning with speakers her older than she is might be an indication that she aims to maintain more traditional norms, though she is still lower in comparison (node 7).

Figure 4.19 shows the ctree for F2 for speakers born after 1987. Men treat TRAP in all the environments as the same, and realise it around a fairly central median of 1700 Hz. The most retracted realisation of TRAP is produced by women in the free environment or before /g/ (node 7). In the other environments, TRAP also retracts (node 6), though TRAP_1 is the least retracted allophone (node 5).

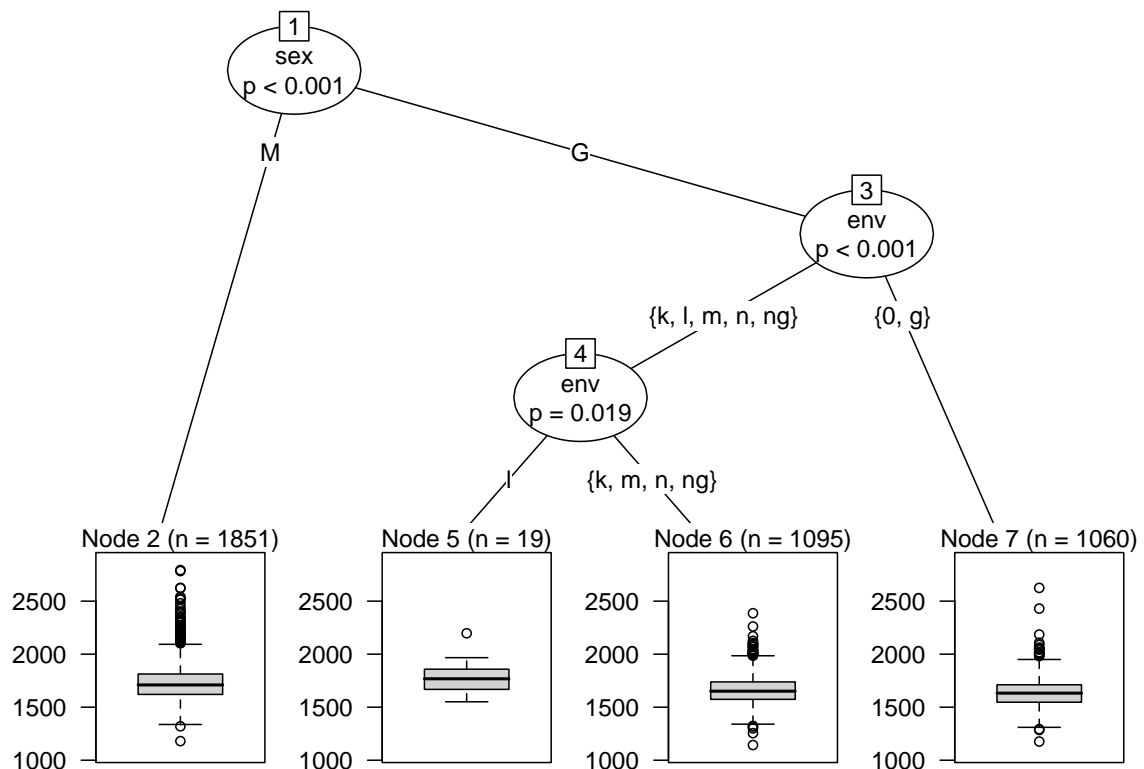


Figure 4.19: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’ and ‘date of birth’ for speakers born after 1987.

Women retract TRAP significantly more than men (compare node 2 with nodes 6 and 7). In these environments, TRAP is retracted towards 1500 Hz, and the fronter variants occur in the expected environments: before nasals and velars. TRAP_1 is the frontest environment for the women in this group, while men make no phonological distinctions. For both genders it is not retracted beyond TRAP_0, which indicates that final /l/ is not a backing environment in this instance.

In summary, there is apparent time evidence for the lowering and retraction of TRAP, both tense and lax subsets. The younger speakers show the lowest and most retracted realisations. Where other Englishes distinguish between tense and lax TRAP regarding both quality and quantity (such as English in New York City), SAfE only distinguishes quantity. This is

evident in the mixed-effect models and in the ctree analyses, where the voicing of the coda does not significantly affect the quality of the vowel.

Compared to the older set of speakers, the younger speakers generally lower the vowel fairly uniformly, with allophonic variation in the speech of young women (figure 4.19). In this sense it would appear that lowering of TRAP is complete for younger speakers. Young women retract TRAP the furthest, followed by young men. The older the speaker, the fronter the vowel is, with the oldest speakers producing the most front variants. This is the third piece of evidence for the Reverse Short Front Vowel Shift: KIT, DRESS and TRAP all lower in SAfE, and they all retract.

Given the shape of traditional (and acoustic) vowel charts, the question of ‘automatic’ versus ‘true’ retraction of the TRAP vowel must be addressed. In other words, is TRAP lowering so much that it automatically retracts when it reaches a certain point, or is the retraction evident in the speech of younger SAfE speakers ‘true’ retraction along the lines of a chain shift? In order to answer this question, figure 4.20 compares the mean realisations of two older women (excluding KayC as she is a continual outlier) to the younger women, who have been shown to retract and lower TRAP in comparison to the older speakers. From this scatterplot it is evident that the retraction of TRAP is not merely a product of TRAP lowering, for the following reason: The young women who retract TRAP most are not necessarily the women who lower TRAP most. The two highest TRAP realisations of the younger speakers are retracted. This shows that it is not only the lowered TRAP realisations that are associated with retraction. Moreover, the most retracted realisation(s) of TRAP are not necessarily the lowest, which supports the notion that, in this variety of English, TRAP retraction is not a consequence of TRAP lowering, but a separate change occurring in SAfE in addition to TRAP lowering.

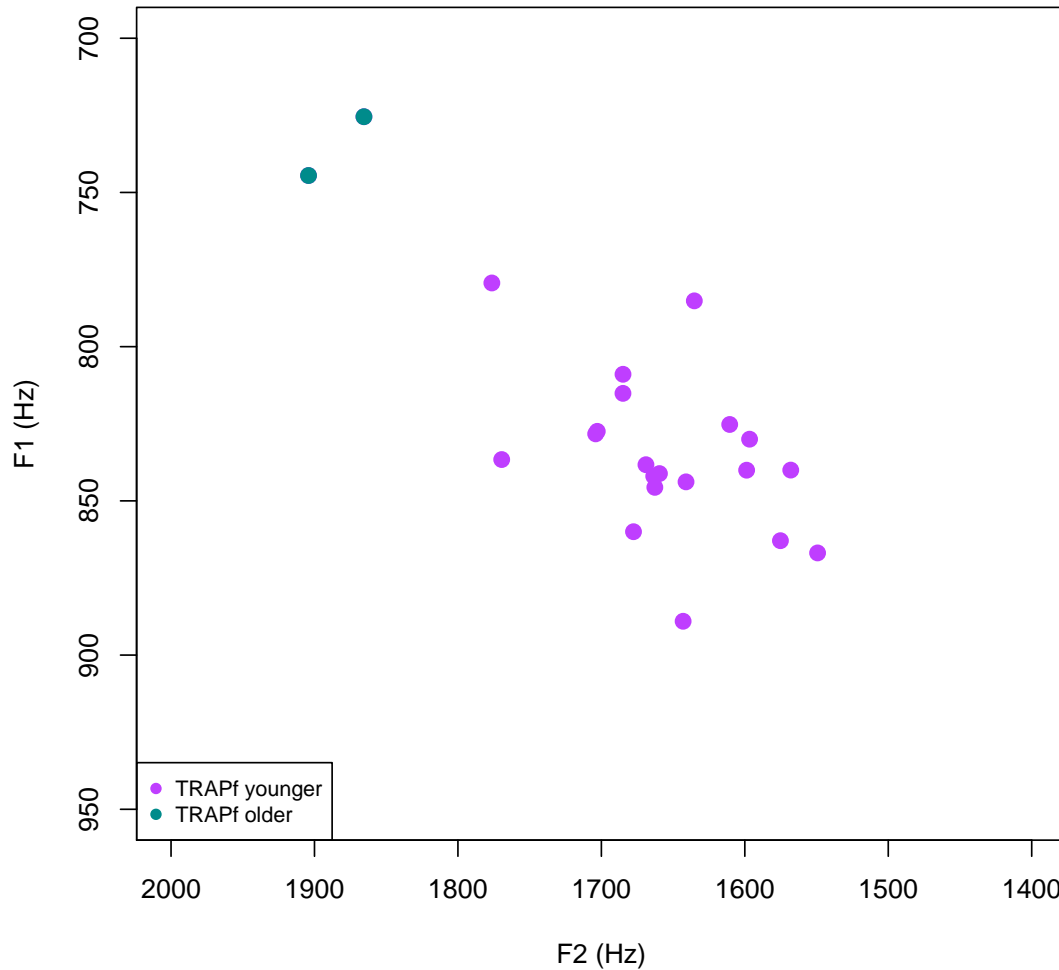


Figure 4.20: Mean values, per speaker, of the TRAP vowel of 20 younger women compared to two older women.

Similar evidence for true TRAP retraction is presented in Kennedy and Grama (2012, outlined in Chapter 1, section 1.4.3). I reproduce their graph below (figure 1.14a) for ease of reference. Though they make no overt claims about TRAP retraction in relation to the limits of the vowel space, the tokens for TRAP in figure 1.14a shows that TRAP can be retracted without having to be extremely low. Much the same as in my data, the most retracted realisations of TRAP are not necessarily the lowest, and vice versa. Furthermore, Kennedy and Grama (2012, 49) find that though there is a significant difference ($p=0.0028$) in F1 between men (841.26 Hz) and women (903.59 Hz); there is no significant difference in the F2 of TRAP (and indeed any other vowel), suggesting that lowering does not precede or necessitate retraction.

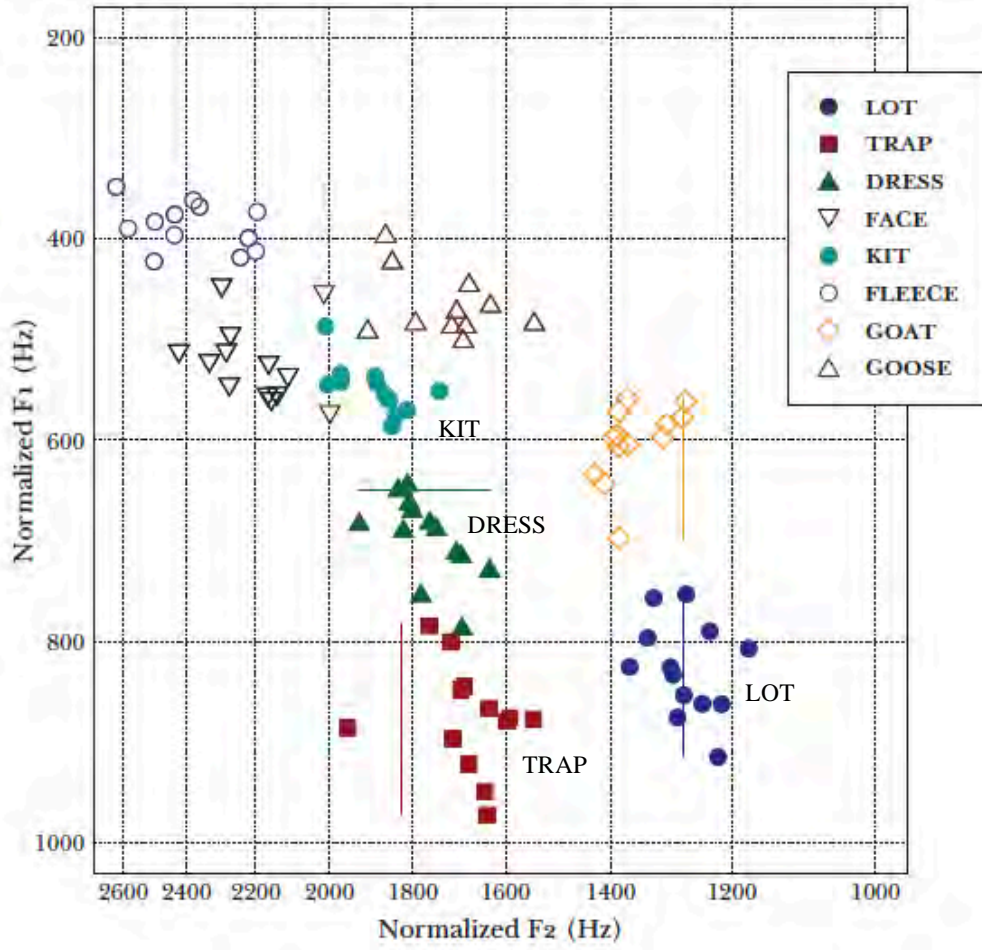


Figure 1.14(a): The California Shift among 13 speakers aged 19 to 29 (Kennedy & Grama 2012, 47).

4.5 THE REVERSE SHORT FRONT VOWEL SHIFT

The previous sections have provided concrete evidence for the existence of a reversal of the Older Vowel Shift. By way of comparison, a brief summary of the Old Shift (fully discussed in Chapter 2, section 2.3.2) is provided. Roughly schematised in figure 4.21, the Older Vowel Shift involved the raising of the short front vowels when compared to Received Pronunciation. The TRAP vowel raised from [æ] to [ɛ]; DRESS raised from [ɛ] to [e]; and KIT centralised to [i̯], with the exception of KIT2.

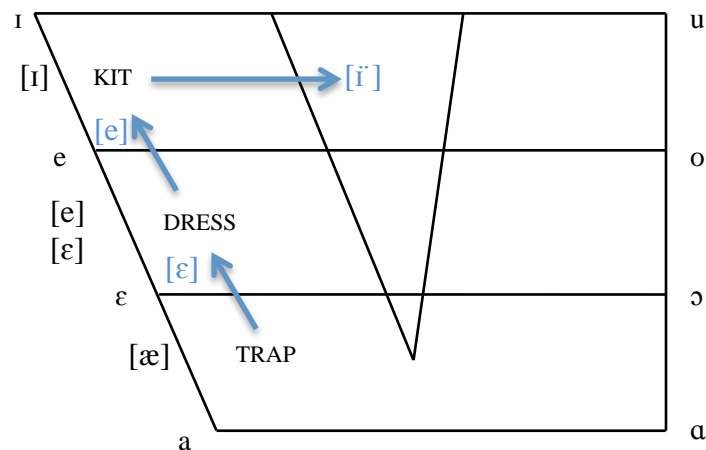


Figure 4.21: Schematic representation of the Older Vowel Shift.

The Reverse Vowel Shift (i.e. the new shift) involves a reversal of the raising characteristic of the Older Vowel Shift (red in figure 4.22). The lowering of the short front vowels is more than a change back to the norms of SAfE before the Older Vowel Shift (for some speakers): the Older Vowel Shift is being reversed beyond the older norms. In this process some, particularly younger speakers, are innovating new norms for SAfE. This is particularly evident in the retraction of KIT, where the fronter allophone, KIT2 retracts (and lowers) towards [i̯], and the already centralised allophones retract (and lower) to something in the region of retracted [ɐ]. The reversal of the Older Vowel Shift is particularly evident in the extreme lowering and retraction of TRAP. Old TRAP raising to [ɛ] is undone, though it does not stop when it reaches the original position [æ]. The old raising is completely reversed by innovative, extreme lowering and retraction to [ä]. Though DRESS shows evidence of lowering, the distance between it and TRAP remains relatively large, suggesting that the raising of old DRESS is undone, though not reversed. In other words, speakers, particularly female speakers, lower DRESS to [ɛ̞], thereby not moving far from the older, higher, prestige

variant. Men lower DRESS further towards [ɛ], though they do not quite reach original [ɛ]. DRESS is further retracted towards [ɞ̥] for young women and [ɛ̥] for young men.

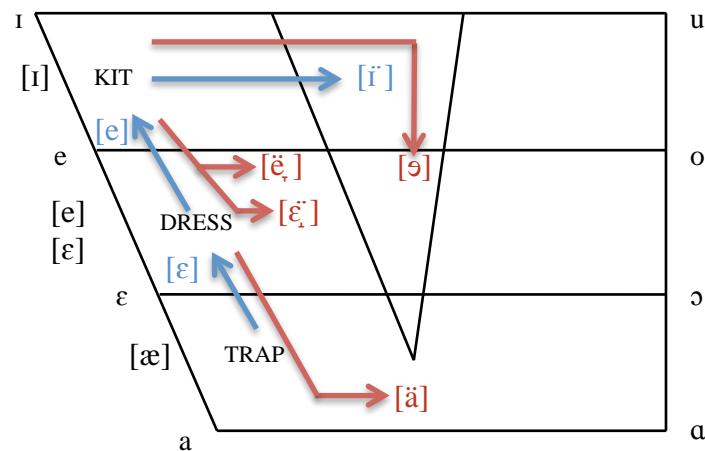


Figure 4.22: Schematic representation of the Reverse Vowel Shift.

Evidence of the Reverse Vowel Shift in the speech of young Capetonians is also provided by the behaviour of a continual outlier, KayC. As mentioned before, she is a 53-year-old woman who, at the time of the interview, was in her first undergraduate year at university. In most cases she is significantly different from others in her age group in her realisations of the short front vowels. These differences are in the direction of the norms of the younger people she is exposed to daily, particularly the lowering and retraction of KIT and TRAP. Though she is still significantly different to her university peers, the direction of the changes in her speech suggests accommodation to the younger norms encompassing the Reverse Vowel Shift.

4.6 OTHER CAPETONIANS

The deracialisation of SAfE is well documented, seminally by Mesthrie (2010). As discussed in Chapter 2, Mesthrie found that young, black middle class South Africans, particularly women, participate in GOOSE fronting to a large extent. Indian and coloured South Africans also participate in this fronting, though neither as cohesively as the black speakers in his sample, nor as far advanced in the process (i.e. they have more centralised GOOSE realisations). In this section I therefore explore to what extent young black middle class Capetonians participate in the Reverse Vowel Shift. As has been the case thus far, the analysis makes use of conditional inference trees and linear mixed-effects modelling. The coding of the data presented here is in accordance with what has been outlined thus far (see also Set B in the list of abbreviations).

Figure 4.23 is a scatterplot of the short front vowels for black and white speakers. Visually, the similarities are striking. The results of the mixed-effect models show that there are no significant differences between the ethnic groups for KIT, DRESS or TRAP.⁶¹ Each short vowel is dealt with individually (sections 4.6.1 - 4.6.3) before I conclude the chapter in section 4.7.

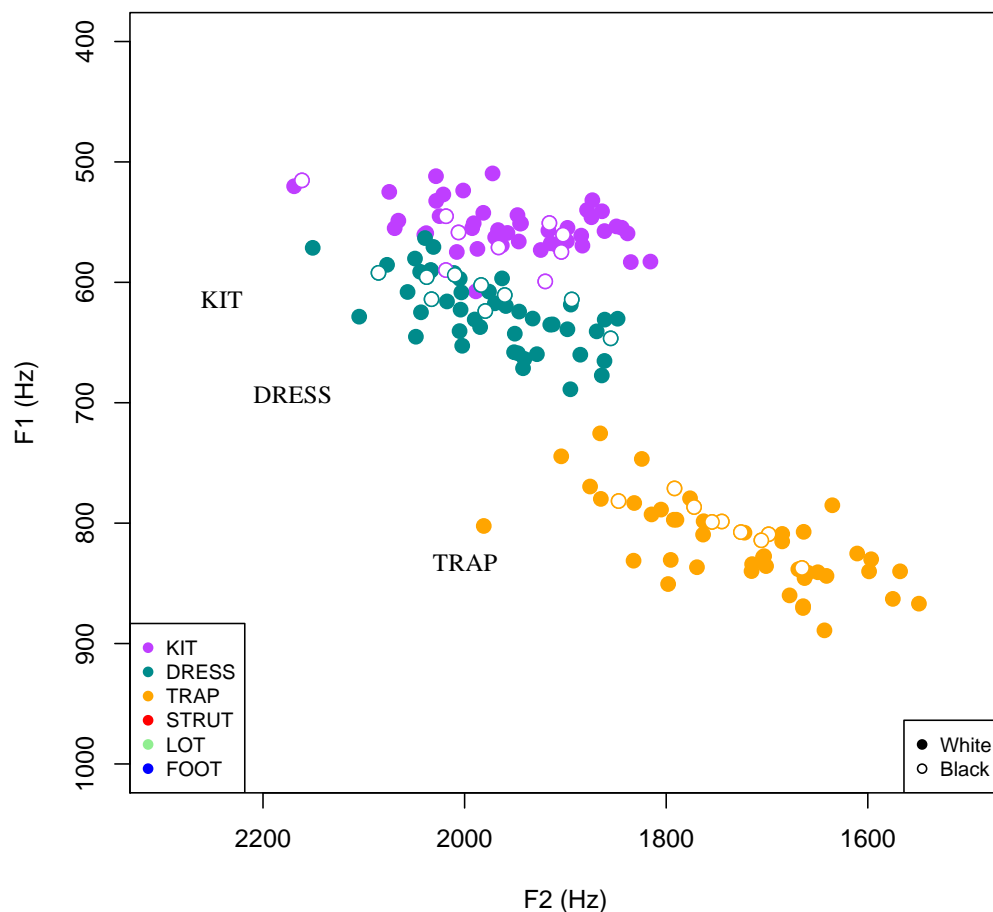


Figure 4.23: Mean values of KIT, DRESS and TRAP per speaker for young white and black middle class Capetonians.

⁶¹ KIT: F1 $p=0.408$; F2 $p=0.209$. DRESS: F1 $p=0.095$; F2 $p=0.297$. TRAP: F1 $p=0.137$; F2 $p=0.551$.

4.6.1 KIT

In addition to the 5520 tokens collected for white speakers, 1454 tokens produced by black speakers are available for analysis. The results of a ctree analysis (figure 4.24) show that the only significant difference between the ethnicities is evident in the KIT2 and KIT6 allophones (node 4): the black speakers realise these allophones slightly lower (closer to 600 mean Hz) than the white speakers their age, significantly different at $p=0.002$ (compare nodes 5 and 6). In addition, there is a greater degree of variation in the realisations of the white speakers, which contributes to the statistical difference. However, given the similarities between the two groups, it follows that KIT lowering is a feature for both black and white speakers of SAfE.

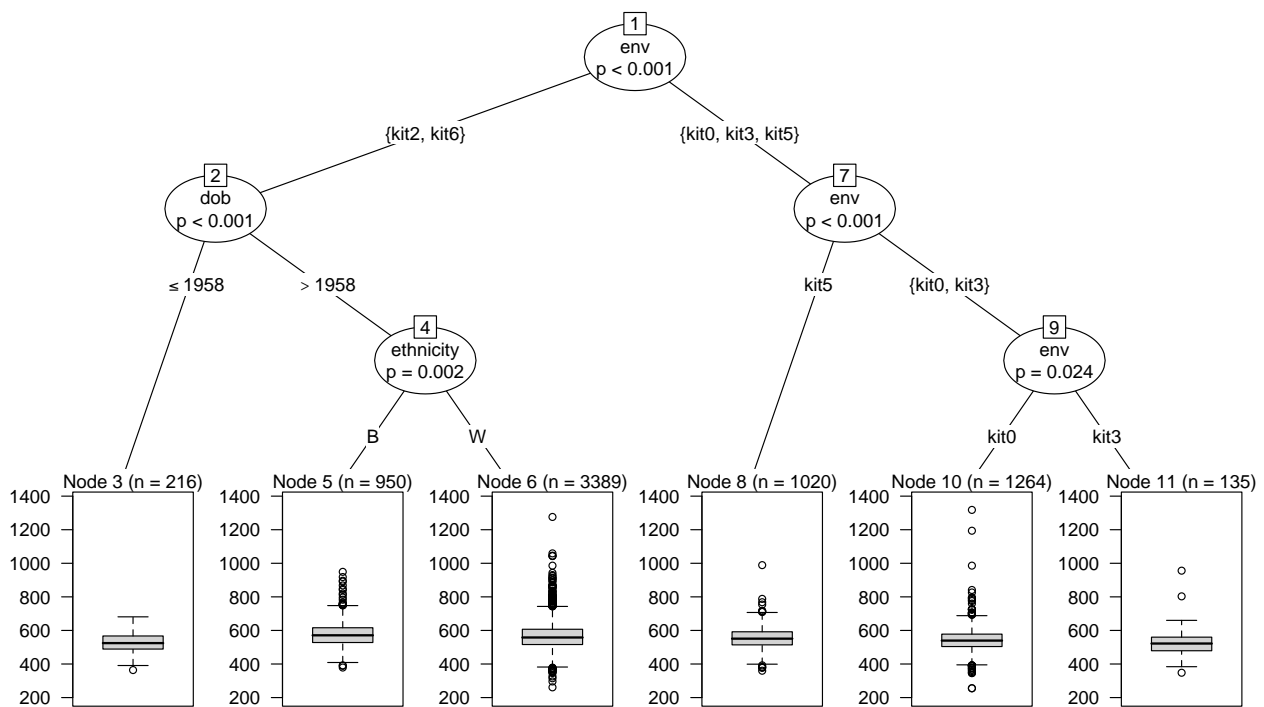


Figure 4.24: Conditional inference tree for KIT: F1 for 'environment', 'gender', 'ethnicity' and 'date of birth'.

The frontness of KIT is more complex, and thus the tree is not provided. KIT2 significantly branches off as separate to the other KIT allophones ($p < 0.001$), suggesting that black speakers participate in the KIT split. An lmer analysis confirms this: KIT2 is significantly ($p < 0.000$) fronter than KIT0. The only other environment significantly different from KIT0 is KIT6 which is significantly retracted ($p = 0.000$), as expected from an environment including final /l/. A conditional inference tree analysis shows that there are no ethnic differences in the realisations of KIT2 regarding F2: black and white speakers realise KIT2 in the same way ($p = 0.398$, lmer). Therefore, the black speakers participate in the retraction of KIT2.

Regarding the centralised allophones of KIT (i.e. ‘other’ KIT), ‘gender’ is the most significant predicting factor (figure 4.25). Within each gender there are ethnic differences that are significant. In the male group there is ethnic differentiation between KIT0 (the free position) and KIT3 (before/after palato-alveolars). Black men (node 4) produce fronter realisations than white men (node 5). The remaining environments show no ethnic differentiation, and are retracted significantly in comparison (node 6).

Within the female group there is no differentiation between KIT0 and KIT3, though there is between KIT5 and KIT6 (node 9). Firstly, black women distinguish between KIT5 and KIT6, where the latter is the most retracted variant (node 12).⁶² Compared to white women, black women do not retract KIT5 as much. In general, the women produce a more centralised version of ‘other’ KIT than the men (comparing medians across all box plots). In this sense, young black women participate in the secondary retraction of KIT, particularly KIT0 and KIT3, though not KIT5. In this way they maintain the KIT split, given that they treat KIT2 in the same way as white speakers. Black men, on the other hand, do not fully participate in this secondary retraction, given their very front realisations of KIT0 and KIT3 (node 4). In fact, a mixed-effect regression shows that black men do not distinguish between KIT2 and KIT0/KIT3 ($p = 0.131$ and $p = 0.216$ respectively). Black men therefore do not participate in the KIT split in the same way as the other speakers: only KIT5/KIT6 are centralised as expected, while the other environments are not. If we ignore the overlaps in nodes 6 and 8, the general frontness of ‘other’ KIT can be hierarchically organised, from frontest to backest, as follows: black men > white men > black women > white women.

⁶² Again, this is expected due to final /l/.

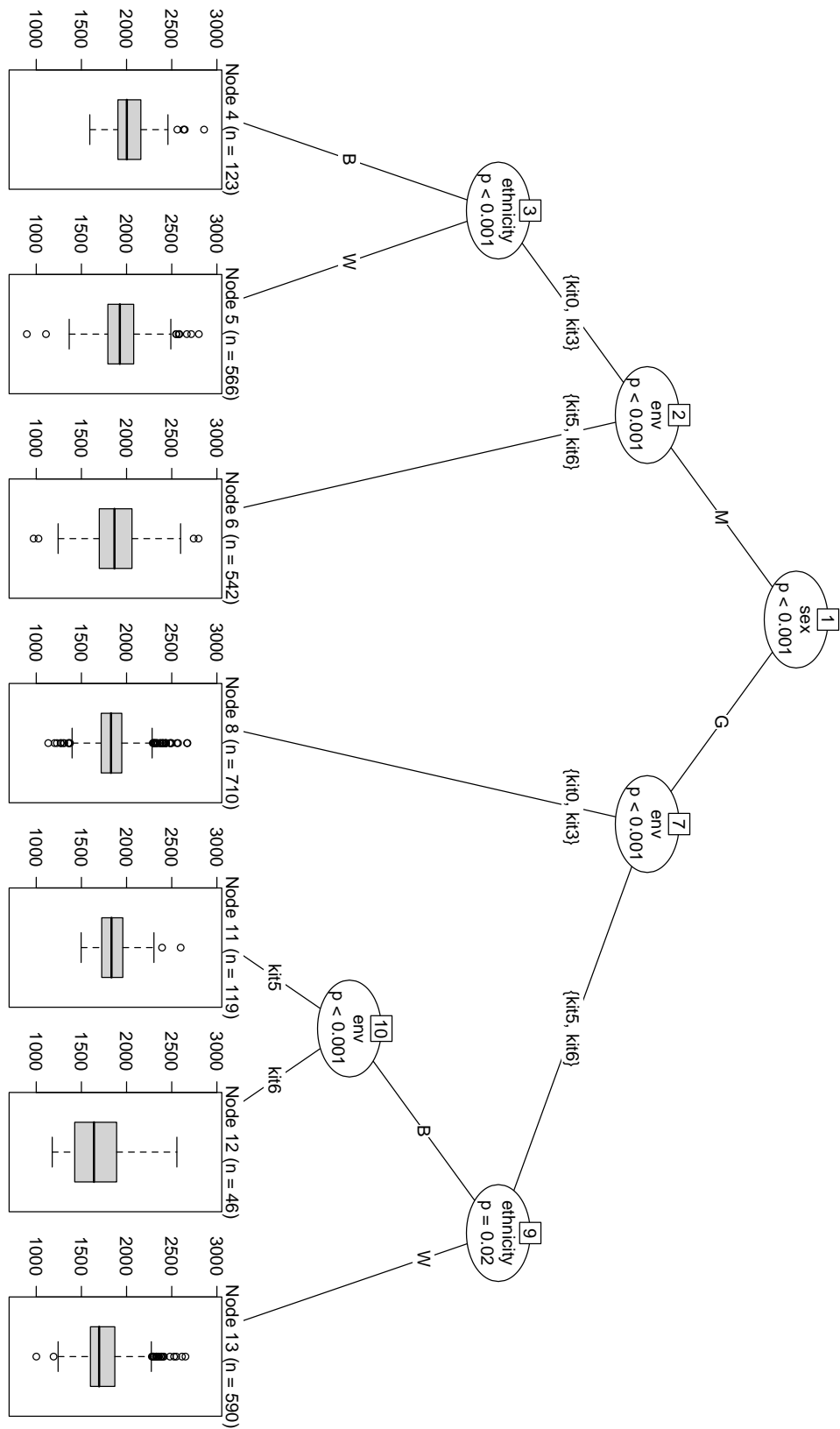


Figure 4.25: Conditional inference tree for 'other' KIT: F2 for 'environment', 'gender', 'ethnicity' and 'date of birth'.

In summary, black Capetonians are similar to white Capetonians in the sample in their realisation of the KIT vowel, though their treatment of it is not exactly the same. The presence of the KIT split in the speech of black women is a feature that has not been explicitly documented before. In addition to producing a KIT split, the black women participate to some extent in secondary backing along with the white speakers. This illustrates the emergence of a new norm for KIT in English spoken by female black middle class South Africans, different to the norms previously associated with black speakers of SAfE, where there was no documented KIT split or KIT centralisation. The lowering of KIT is also a new feature used by both genders, and KIT splitting and centralisation may be a change led by women.

4.6.2 DRESS

The results of a ctree analysis for DRESS reveals that all speakers realise DRESS before /l/ significantly differently to DRESS in the other environments. Height is significantly variable as is evident in figure 4.26, where black speakers realise DRESS_l lower than white speakers.

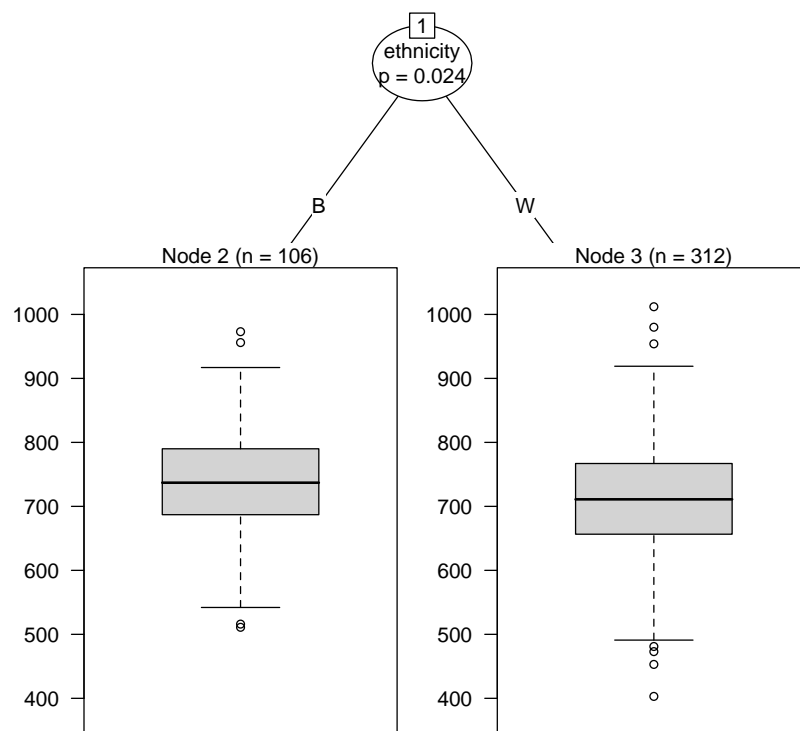


Figure 4.26: Conditional inference tree for DRESS_l: F1 for 'gender', 'ethnicity' and 'date of birth'.

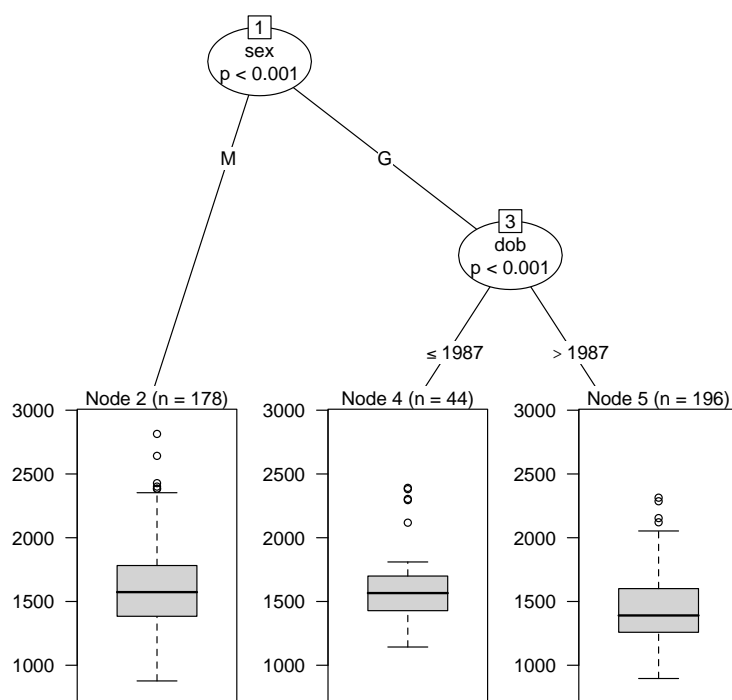


Figure 4.27: Conditional inference tree for DRESS_1: F2 for ‘gender’, ‘ethnicity’ and ‘date of birth’.

Figure 4.27 shows that there were no ethnic differences for the F2 of DRESS_1, so retraction before final /l/ is uniformly carried out by both ethnicities. Importantly, all black women participate in the secondary retraction of DRESS_1 along with white women born after 1987 (node 5).

The remaining DRESS allophones are also subject to variation, with ‘gender’ being the most significant predictor. For women (figure 4.28), speakers under the age of 24 differentiate height based on ethnicity (node 7).⁶³ In general, the realisations of DRESS by black women are higher than those produced by white women, especially in nodes 10-12 in comparison with node 15. The only overlap in height between the two groups is evident in comparing node 14 with node 12.⁶⁴ Black women lower DRESS_m,n as much as white women lower DRESS_0,g. Despite this similarity, the generally lower qualities of DRESS as realised by the white women suggests that black women do not lower DRESS as much, or in the same way. The spread of the outliers in nodes 14 and 15, however, suggest a lowering tendency, which may reach significance in the future.

⁶³ There are no black women over 24 years of age in the sample.

⁶⁴ Node 11 represents one speaker, CindyH, who is often an outlier.

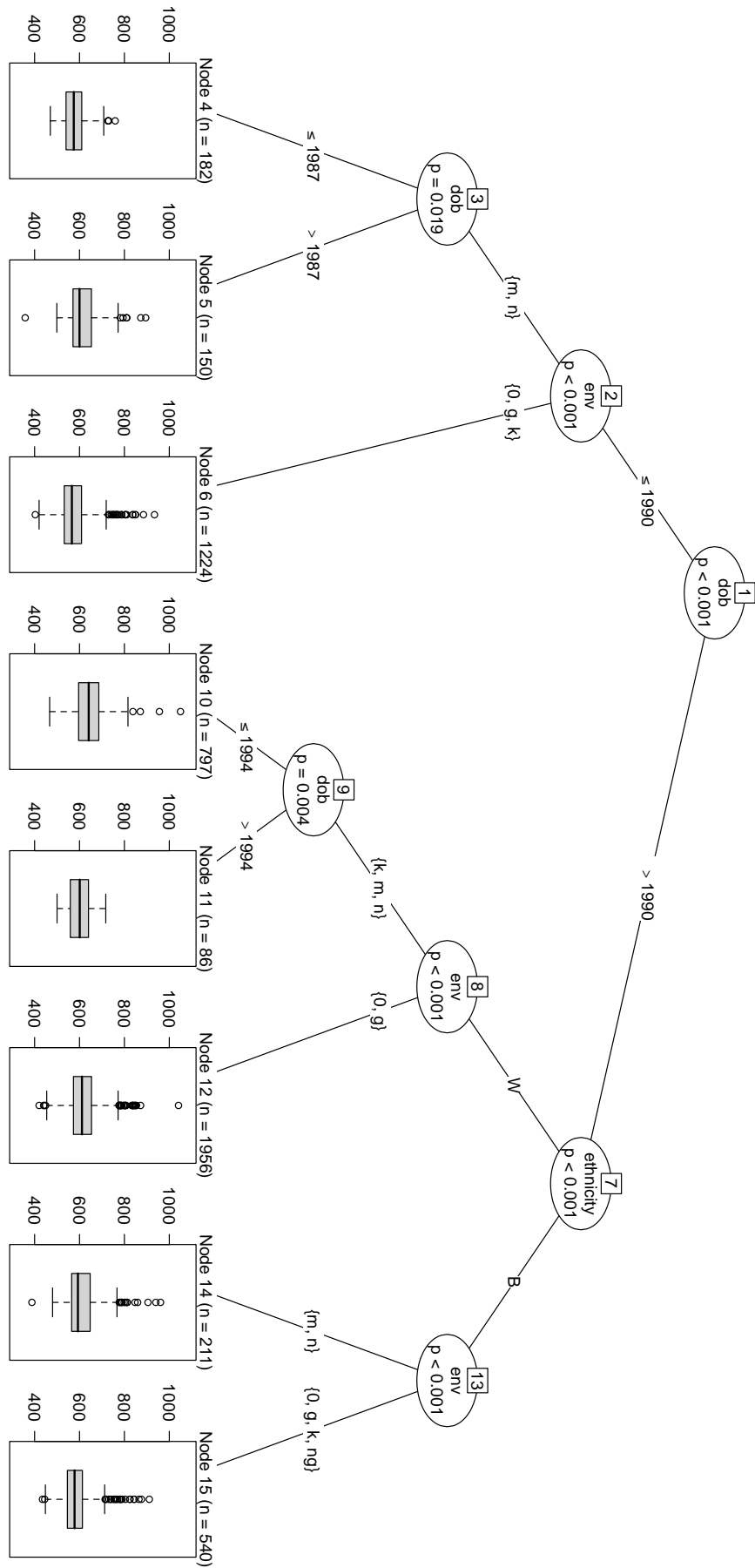


Figure 4.28: Conditional inference tree for ‘other’ DRESS for women: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.

The results for the men are presented in figure 4.29. In contrast to the results for the women, ‘ethnicity’ is the most significant predictor of the height of DRESS.

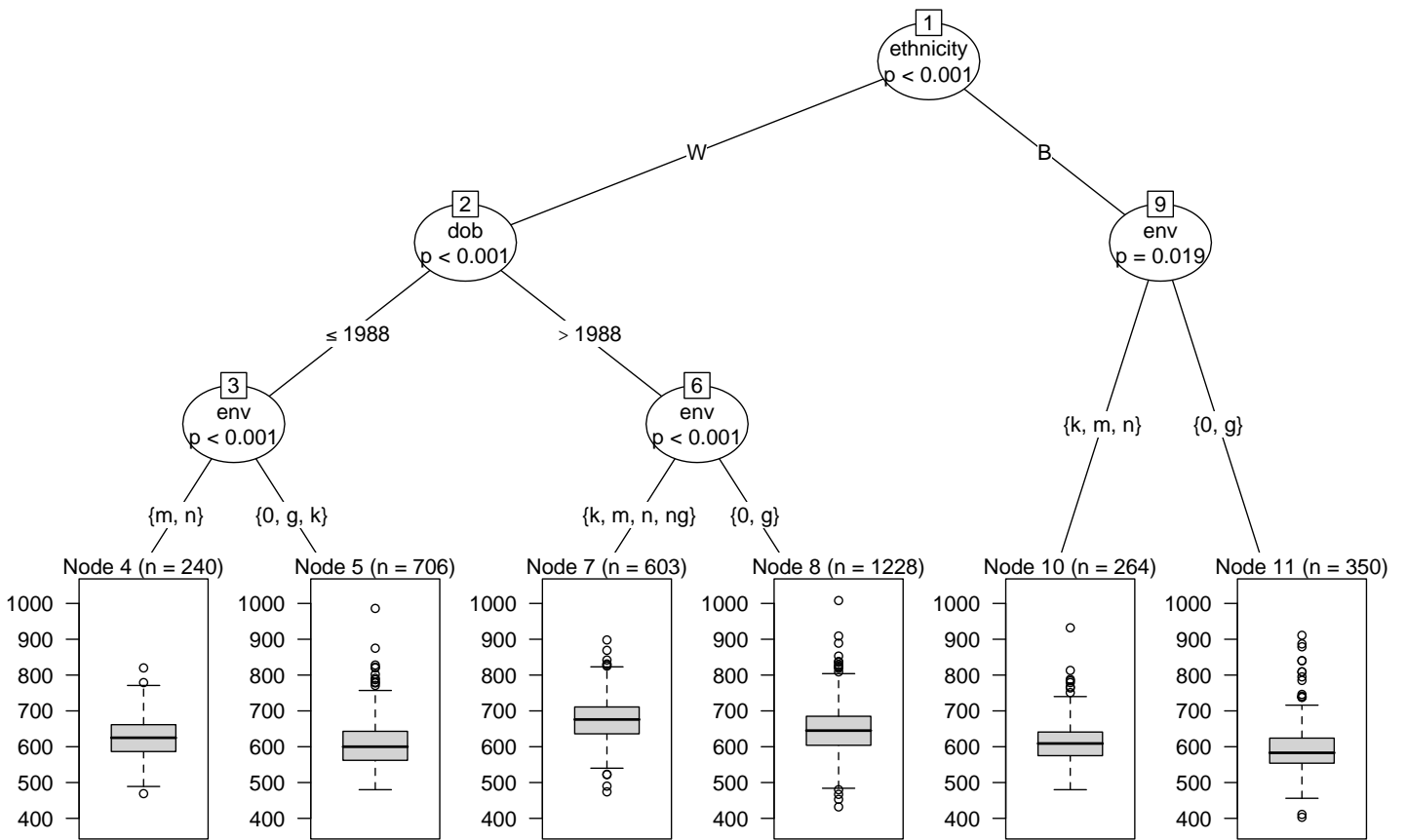


Figure 4.29: Conditional inference tree for ‘other’ DRESS for men: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.

The black men generally have much higher realisations than white men their age, particularly evident when comparing the boxplots contained in nodes 6 and 9. The black speakers in this sample overlap more with slightly older white speakers (node 3), though their realisations in general are still higher.

When compared to black women (figure 4.28, node 13), the realisations of the black men are similar in two ways: (1) both centre around 600 Hz, and (2) both have lowering tendencies. A mixed-effects model shows no difference in the height of DRESS based on ‘gender’ for black speakers ($p=0.562$). In addition to the generally lower realisations by white speakers, this shows that black speakers do not fully participate in DRESS lowering, despite certain similarities and a tendency to lower the vowel.

Regarding frontness, a ctree analysis groups black speakers with white speakers born between 1981 and 1989 (node 5, figure 4.30). After this, the black speakers are significantly different ($p < 0.023$) and branch off separately. Only one black speaker is represented here: ThukileM born in 1988. His median realisation of DRESS is just over 2000 Hz, and none of the independent variables significantly predict F2 (node 13).

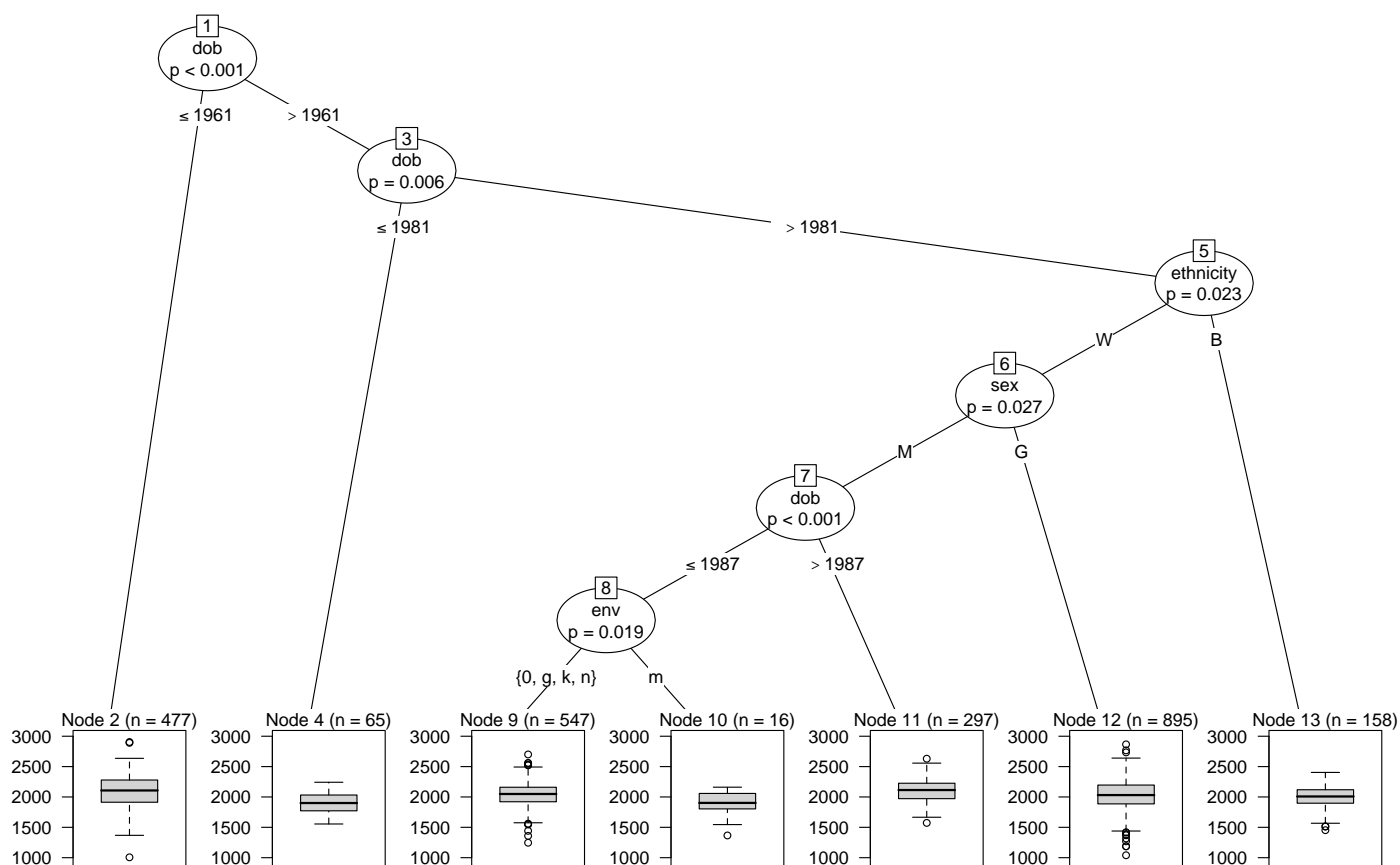


Figure 4.30: Conditional inference tree for 'other' DRESS for speakers born in or before 1989: F2 for 'environment', 'gender', 'ethnicity' and 'date of birth'.

ThukileM realises DRESS in a very similar way to younger white women (compare nodes 12 and 13), retracting it more than younger white men. The rest of the black speakers (born between 1991 and 1996) are grouped with the white speakers born after 1989, as seen in figure 4.30. In this age group, 'ethnicity' is the strongest predictor of DRESS frontness, with black speakers consistently realising the vowel as more front than the white speakers. Though the spread of the tokens for ThukileM (node 13, figure 4.30) is not as big as the other speakers in node 2 (figure 4.31), the median value is stable around 2000 Hz. Generally, therefore, black speakers do not retract DRESS to the same degree as white speakers, who generally retract beyond 2000 Hz (median).

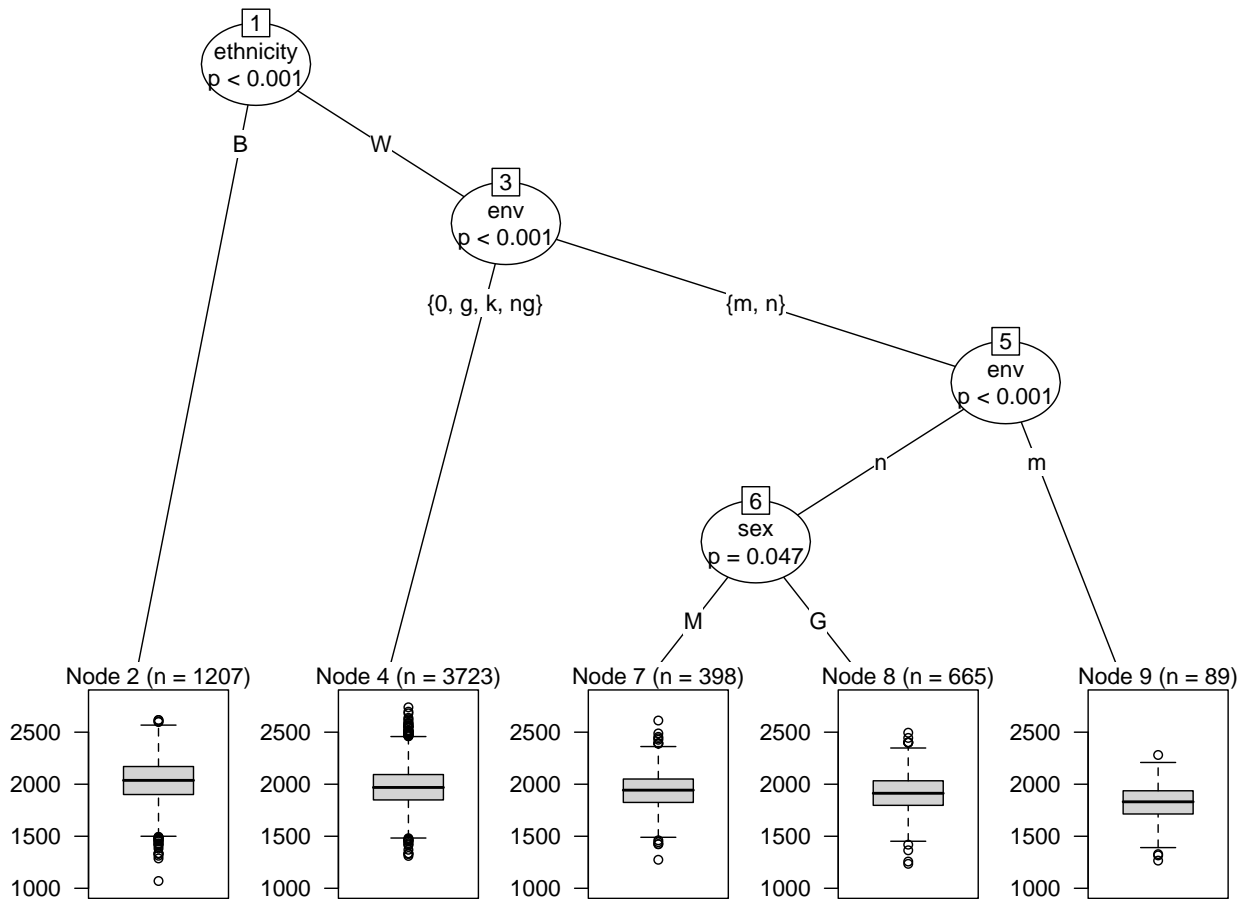


Figure 4.31: Conditional inference tree for ‘other’ DRESS for speakers born after 1989: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’.

The descriptions in Chapter 2 ascribe general coherence to [ɛ] as the variable for DRESS in acrolectal BSAfE, and [e] or something slightly lower for white speakers of SAfE. Wilmot (2014) finds the lowering of DRESS to [ɛ] in the speech of the white private school girls, which is not significantly different to the realisations by the black private school girls. Wilmot (2014, 333) concludes that this is evidence of accommodation to a prestige variable. The black speakers from the former model C schools, however, have a higher realisation of DRESS, which Wilmot (2014, 334) concludes is accommodation to the previous, raised norm for DRESS as part of the Older Vowel Shift.

In contrast to Wilmot’s findings, the data presented above shows consistent differences between the black and white speakers in the sample. Importantly, the white speakers retract and lower DRESS, while the black speakers largely do not in comparison, maintaining an [e]-like realisation. It is possible that schooling is important: Wilmot’s results show that lowered DRESS is a feature for the girls in private school, whereas all the black speakers in my sample attended former model C schools, which is where Wilmot finds higher DRESS realisations. Furthermore, Wilmot’s sample drew from two boarding schools, which could also be a

relevant difference between her sample and mine given the high degree of accommodation likely in a boarding school setting. Age and region may be additional factors, in that my data has slightly older speakers from Cape Town, compared to her schoolgirls from the Eastern Cape. It may also be that white speakers lower DRESS more as they are active participants in the Reverse Vowel Shift, while the black speakers do not have this shift driving the lowering of DRESS. Importantly, the black speakers in my sample do not pattern with the older (1958-1961) speakers, and so do not use norms associated with the Older Vowel Shift such as [e̞] or [ɪ]. Instead, they maintain the use of [e] as indicated by Wilmot.

4.6.3 TRAP

After various exclusions, 1188 tokens of TRAP for black speakers were available for analysis in addition to the 5413 tokens for white speakers. Figure 4.32 shows the ctree results for TRAP for F1 for speakers born after 1987.⁶⁵ ‘Ethnicity’ is selected as the most prevalent predictor, and each ethnicity treats TRAP differently regarding ‘gender’, ‘environment’ and ‘date of birth’.

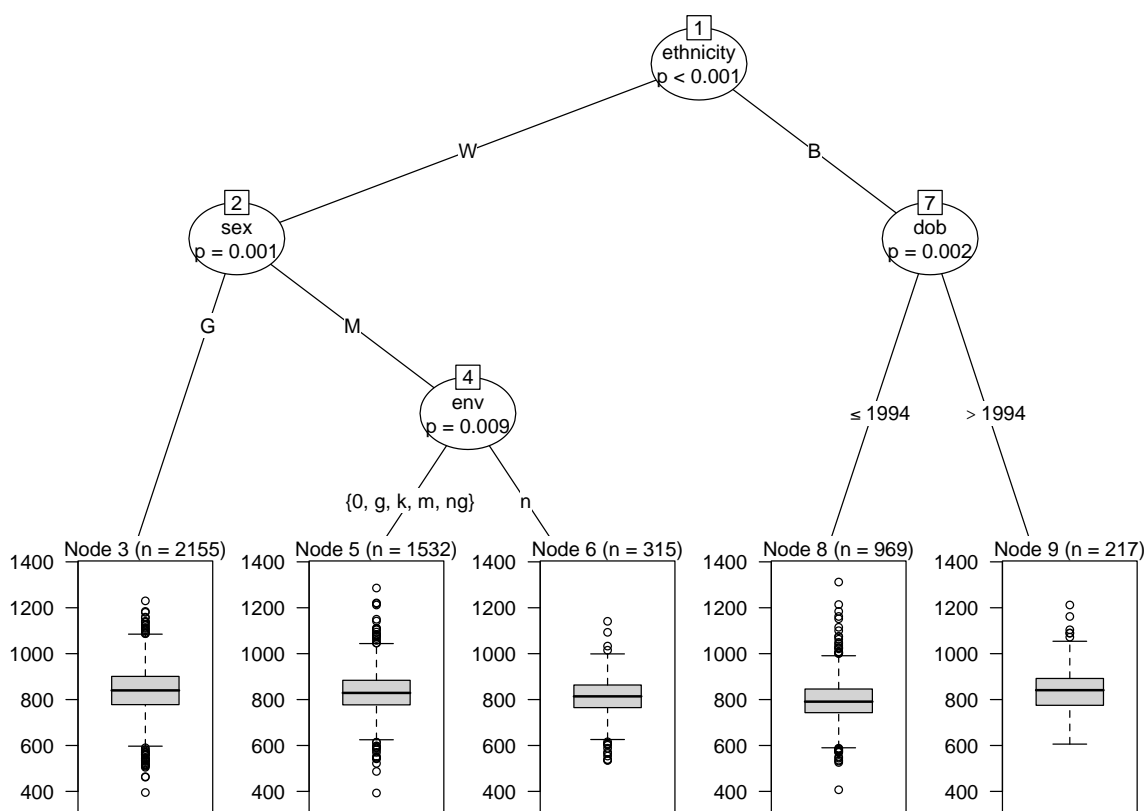


Figure 4.32: Conditional inference tree for TRAP: F1 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’ (>1987).

⁶⁵ The initial tree drew a distinction based on this date of birth.

Young white women lower TRAP more than young white men, and in this sense lead the change (as discussed in section 4.4). Black speakers born between 1987 and 1994 have higher realisations of TRAP (node 8). These speakers include everyone except AndiswaM who was born in 1996 and who is the sole speaker represented by node 9. She is the only black speaker who realises DRESS in a similar way to young white women (node 3), and generally lowers it slightly more than young white men. If we consider her as an outlier, black speakers in general do not lower TRAP as much as is the case for users of the Reverse Vowel Shift, though node 8 shows a lowering tendency.

The degree of retraction is predicted in a more complex way (figure 4.33). ‘Gender’ is the most important predictor, which shows that men and women are significantly different in their realisations of TRAP along F2. Regarding men (node 2), white men generally retract TRAP more than black men (compare nodes 5, 7 and 8). The black speakers are significantly different from one another based on age: node 5 represents two speakers, and nodes 6 and 7 each represent one speaker. In contrast to the other black men, ThembaM (node 6) retracts TRAP in a very similar way to white men, evident not only in the retracted median, but also in the general backward spread of the whiskers. The majority of black women (node 11) generally retract TRAP more than black men, though not as much as white women (nodes 14 and 15). AndiswaM (node 12) is again the only outlier, and retracts TRAP in much the same way as white women. In this sense it would appear that black women generally participate in TRAP retraction, though most of them do not retract as far as the white female speakers.

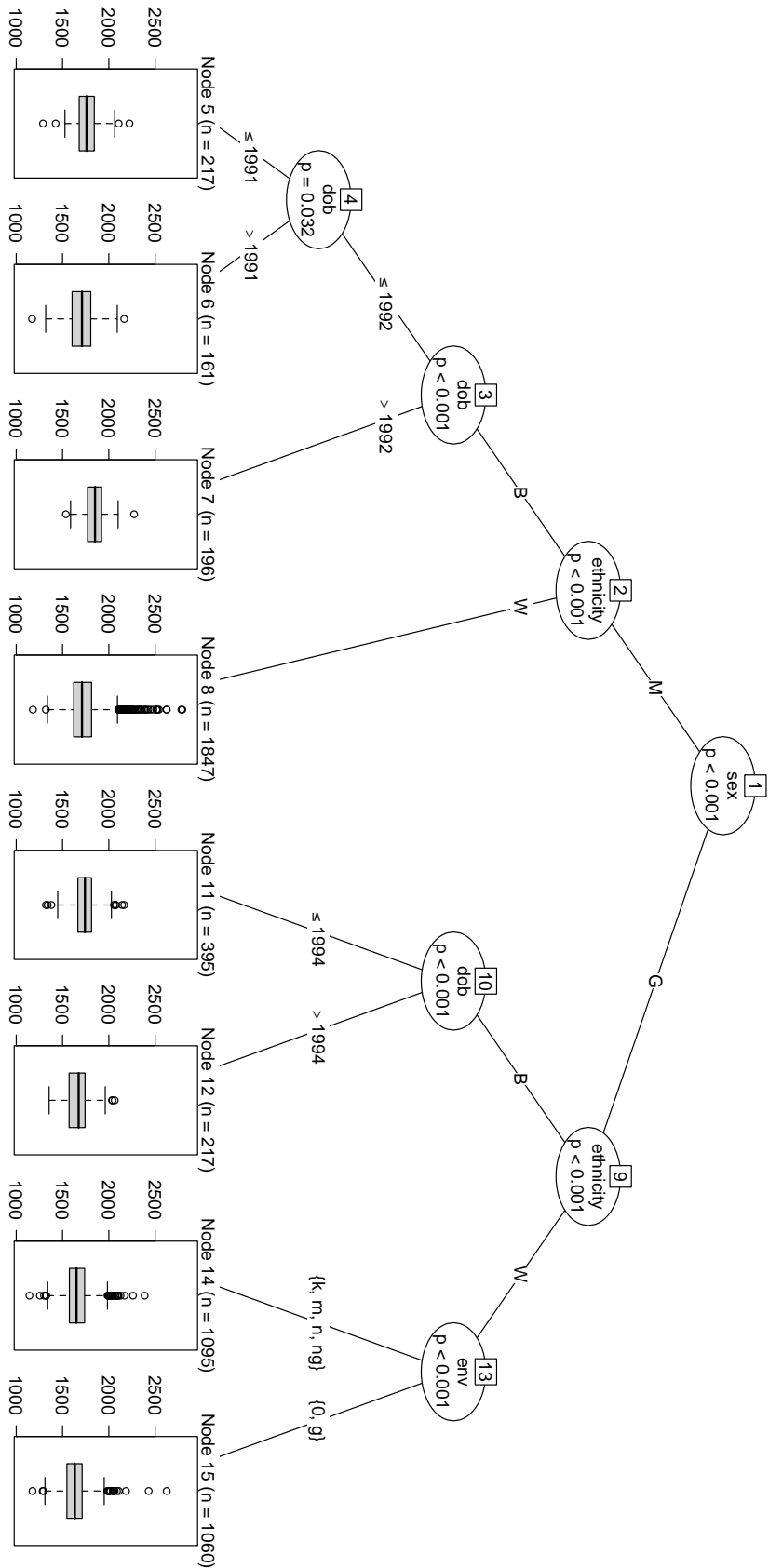


Figure 4.33: Conditional inference tree for TRAP: F2 for ‘environment’, ‘gender’, ‘ethnicity’ and ‘date of birth’ (>1987).

To summarise, black speakers do not uniformly participate in the lowering of TRAP, though they do participate in TRAP retraction to a certain extent. Van Rooy (2004) described TRAP in BSAfE as ranging between [ɛ] and [æ] in acrolectal speech. My data suggests that the black speakers (of an admittedly different educational and background/class) in general prefer the lower [æ] in terms of height, and that the women retract the vowel to something in the region of [æ]. Wilmot's (2014) study is largely supported by these findings. The black private school speakers in her sample realise the vowel as [æ], where the white speakers retract to [æ] or [a]. The black former model C girls in her sample maintain [ɛ] or [æ], whereas the white speakers show slightly retracted [æ]. In my sample, therefore, black and white speakers, especially women, retract TRAP. The lowering and retraction of TRAP to [ä], however is only found for the young white speakers. Young black speakers prefer [æ] (men) or [æ] (women).

4.7 CONCLUSION

Black men retain more raised variants (particularly for TRAP), while black women produce vowels more (but not entirely) similar to the lowered variants of white women. The presence and maintenance of the KIT split in black women's speech in particular shows that KIT is the vowel where the most similarities lie. Black women participate in TRAP retraction to some extent: only one of the five women (AndiswaM) lowers and retracts this vowel to [ä] in the same way as the young white women in the sample. The remaining four women do not lower the vowel beyond [æ], though they retract slightly to [æ]. The DRESS vowel is where the most differences lie (excluding DRESS_1): black speakers neither retract nor lower DRESS in a way similar to the white speakers.

Therefore, the black women in the sample only participate in the following trends common in the speech of the white speakers:

1. KIT split.
2. Retraction of KIT2, secondary retraction of KIT 'other'.
3. Secondary backing of DRESS_1.
4. Partial retraction of TRAP.

Given their participation in some but not all features, black women are not full participants in the Reverse Vowel Shift currently on-going in Capetonian SAfE. These four similarities or partial similarities between the two ethnicities are signs of convergence: where the two ethnic

varieties were traditionally very different, this is no longer the case. This finding supports the thrust of the work by Da Silva (2007) and Mesthrie (2010), insofar as English in South Africa, particularly with regards black and white speakers, is being 'deracialised'. Some of the speakers in these two as well as the present study may even be said to speak a crossover variety, more similar to the norms of white speakers of English (prior to the Reverse Vowel Shift). Accordingly, this raises interesting sociological questions about convergence and new divergences, to which I return in Chapter 6, section 6.4.

CHAPTER 5: RESULTS: OTHER VOWELS AND MERGERS

This chapter presents the analysis and results of the remaining vowels of relevance to the Reverse Vowel Shift. Given the fact that the norms for the black speakers are not concretely indicative of the Reverse Vowel Shift, the analysis in this chapter focuses on the white speakers as users of the new shift. In particular, this chapter considers the relative position of STRUT in section 5.1 and FOOT in section 5.2. A brief analysis of LOT and THOUGHT is provided in section 5.3. The analyses of potential mergers in South African English, namely KIT~DRESS and TRAP~STRUT, are presented in sections 5.5 and 5.6 respectively. Section 5.7 provides an overview of the Reverse Vowel Shift, taking these analyses into account.

5.1 THE STRUT /ʌ/ VOWEL

The analytical approach to STRUT, traditionally a back vowel, was slightly different to the vowels discussed thus far. In the aural observations informing this study, I encountered realisations of STRUT very similar to TRAP (i.e. potentially merged), though I could not ascertain with certainty whether the preceding or following segment was responsible. I thus coded the data for both, following the general principles of environment selection employed for DRESS and TRAP with the addition of r_, as shown in table 5.1. Though STRUT occurs after /g/, such as the word *gut*, this condition was not present in my dataset.

Environment	n	Example
0	1215	<i>study</i>
k (_g)	52	<i>suck</i> <i>cup</i> <i>(bug)</i>
l	311	<i>love</i> <i>culture</i> ⁶⁶
_m/n _	1214	<i>come</i> <i>muscle</i> <i>fun</i> <i>nuts</i>
ŋ	108	<i>young</i>
r_	258	<i>brother</i>
TOTAL:	3158	

Table 5.1: Number of tokens of STRUT by environment.

⁶⁶ I coded ‘culture’ as occurring ‘before /l/’ since the quality of STRUT in this instance is not the same as when it is preceded by /k/ - it is retracted. This was a principal decision based on the well-known backing influence of /l/.

The descriptions outlined in Chapter 2, section 2.2.1 indicate that STRUT in SAfE is generally fronted, realised in the low central vowel space, ranging between [ä] and [a̠]. In certain idiolects, STRUT may retract before /l/ towards [ɔ] (Lanham 1967, 82). Bekker (2009) describes STRUT as a fairly fronted vowel, which overlaps to a large extent with TRAP (also found by Bekker & Eley 2007).

Given the many environments, figure 5.1 shows only the standard deviation ellipses for the different environments. The mixed-effects results show that STRUT is significantly retracted before /l/ ($p=0.011$). Though STRUT tokens before velar plosives are fronter than STRUT_0, they are not significantly so ($p=0.62$ and $p=0.71$ respectively).

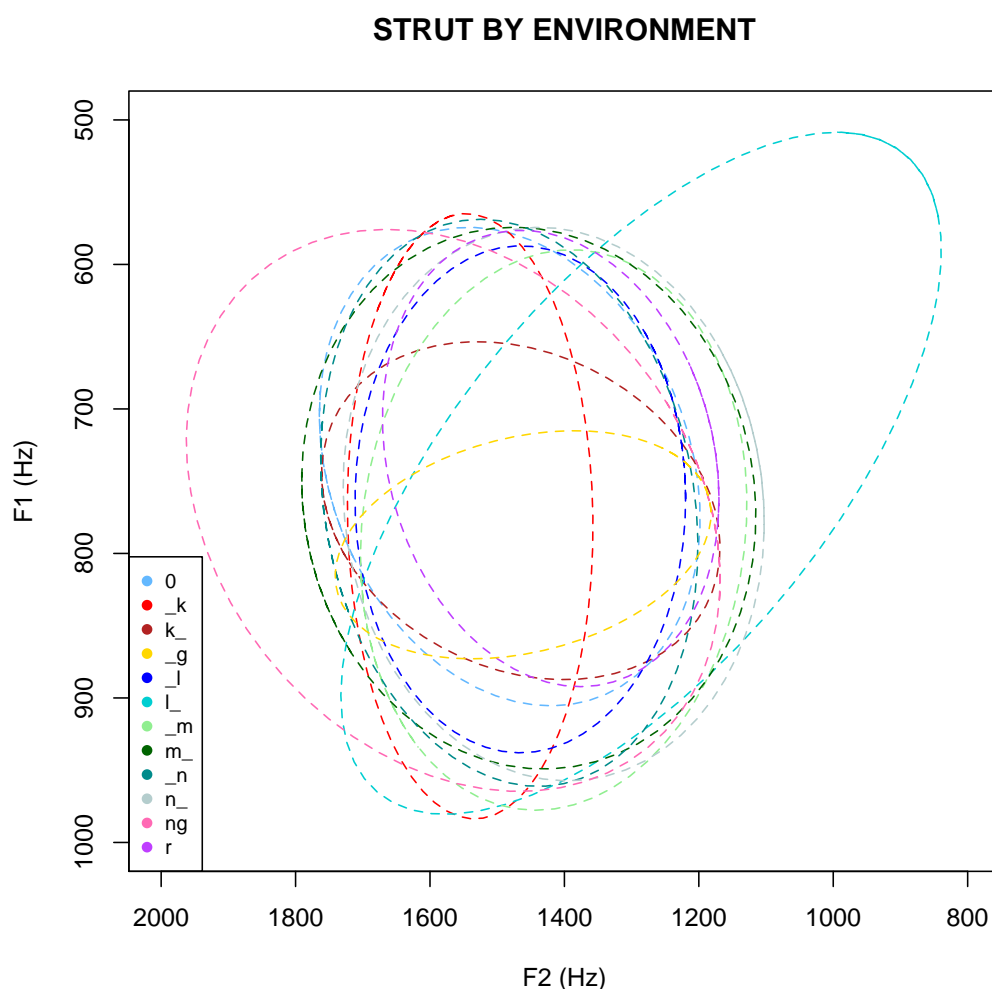


Figure 5.1: Standard deviation (x2) ellipses of the subclasses of the STRUT vowel per environment.

STRUT significantly retracts when it occurs before /m/ ($p=0.027$); nasals are a known retracting environment for back vowels (Labov 1994, 275). Most of the environments have lower realisations than STRUT_0, except for STRUT_1 and r_STRUT, though these differences are not significant.⁶⁷ The only environments significantly lower are m_STRUT ($p=0.000$), n_STRUT ($p=0.010$), and to a lesser extent k_STRUT ($p=0.070$). Velar environments that follow STRUT do not lead to significant lowering, though these realisations are generally lower.

The conditional inference tree for F1 (in figure 5.2) selects ‘environment’ and ‘date of birth’ as significant predictors. In the coding of the environments, ‘a’ indicates that STRUT occurs after the consonant, and ‘b’ indicates that STRUT occurs before the consonant. Therefore ‘ak’ indicates STRUT when it occurs after /k/, as in the word *cup*. Similarly, ‘bk’ indicates STRUT occurring before /k/, in words such as *suck*. Most of the environments pattern together, and have median realisations close to 800 Hz (node 2). In STRUT0, STRUT_1 and r_STRUT (node 3), the vowel has been raising over time, with people over 31 years of age (i.e. the older group including one younger speaker) having lower realisations than people aged 31 or younger (the younger group).

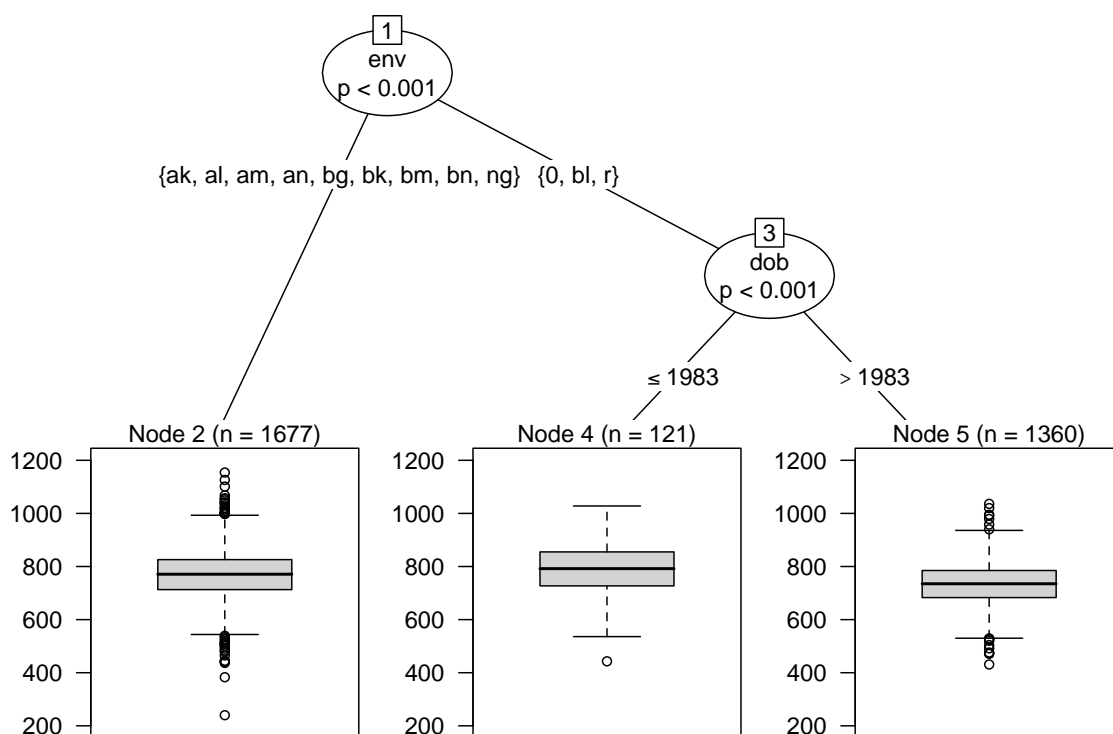


Figure 5.2: Conditional inference tree for STRUT: F1 for ‘environment’, ‘gender’ and ‘date of birth’.

⁶⁷ $p=0.75$ and $p=0.14$ respectively.

The fixed-effect factors have a much stronger impact on the frontness of STRUT, evident in figure 5.3. STRUT is retracted by all speakers when it occurs after /m/, before /l/, /n/ and after /r/ (node 11). These are all known to condition backing (except for preceding /r/) (Labov 1994, 275). The most front realisations of STRUT occur, for all speakers, after /k/ and before /ŋ/ (node 2), neither of which are known fronting environments. The lack of phonetic patterning, and the fact that the most front and most back realisations are significantly different from one another ($p < 0.001$), suggests that the vowel is changing (since allophonic differentiation for STRUT has not featured much in the literature, except for the possibility of backing before /l/).

The remaining environments (node 2) are conditioned by age and gender. All speakers born after 1990 (i.e. under the age of 24) are uniform in their realisations of STRUT (node 10). Though the spread along F2 is generally more advanced than the retracted variants (compare node 10 to node 11), the medians are similar at about 1400 Hz. Speakers under the age of 24 therefore retract STRUT (unless it is preceded by /k/ or followed by /ŋ/).

Speakers over the age of 24 (i.e. born in or before 1990) condition STRUT differently, based on their gender. Women produce realisations of STRUT that are not as retracted as (a) speakers under 24 or (b) the most retracted variants in node 11, though they are significantly backer than the most front realisations in node 3. Men between 27 and 60 years of age (node 8) have very similar realisations to these women, though younger men (19-26 years of age) retract this vowel significantly in comparison (node 9).

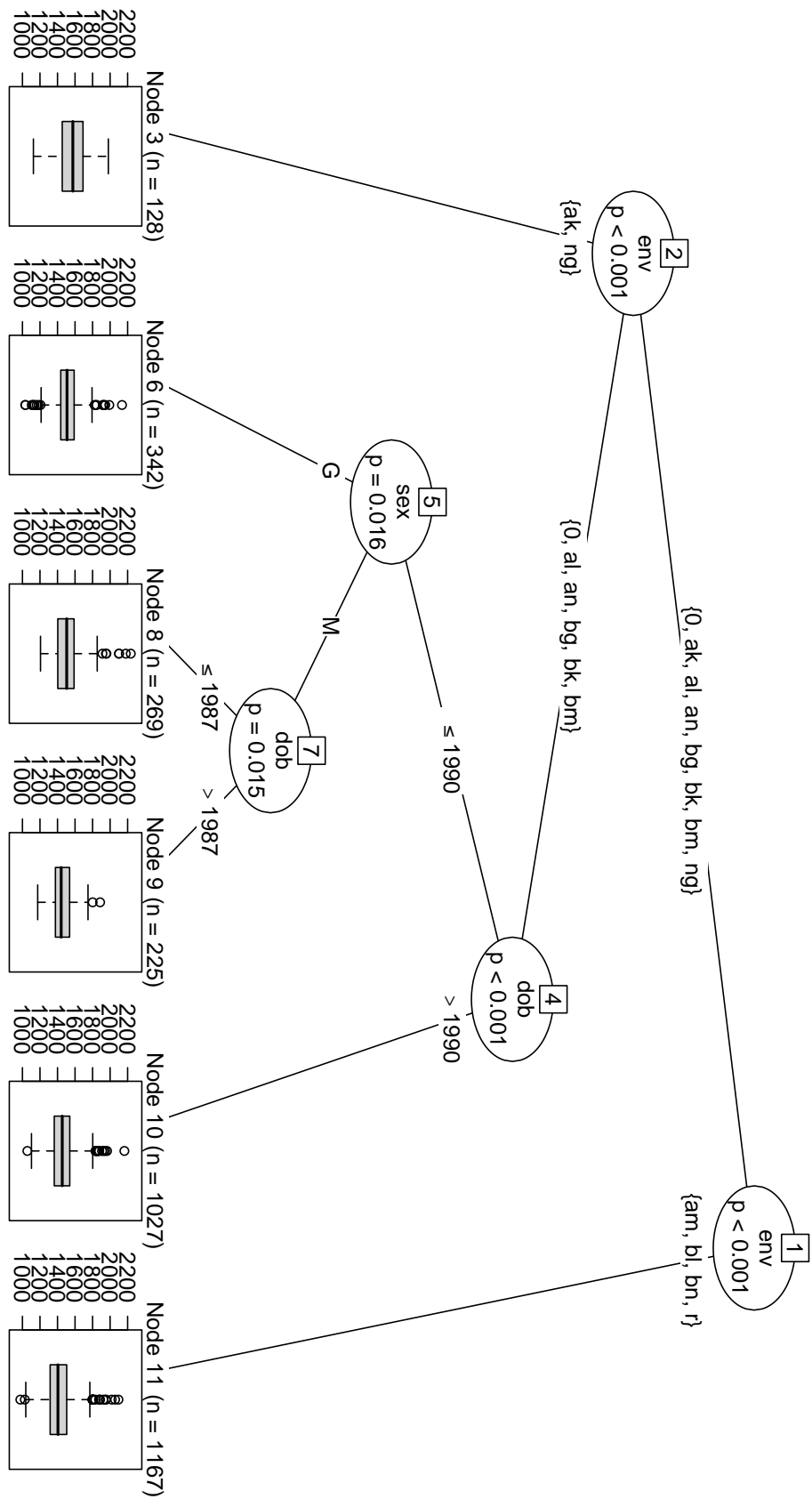


Figure 5.3: Conditional inference tree for STRUT: F2 for 'environment', 'gender' and 'date of birth'.

From these apparent time results it appears that STRUT in is in the process of raising and retracting. All young speakers have higher realisations of STRUT than the older speakers. Retraction is more complicated. STRUT_k and STRUT_ŋ remain stable as central vowels. Retraction is evident in the remaining environments, though the speakers are not uniform in their retractions. Women aged between 24 and 56, and men aged between 27 and 60 retract the least. Men between 24 and 26 years of age retract more than these two groups, but less than speakers between 19 and 24 years of age. In general, however, the younger the speaker, the further retracted STRUT is, though not as far as when it occurs after /m, r/, or before the backing environments /l, n/.

These results confirm what is broadly postulated earlier in the chapter: there is no TRAP~STRUT merger. The changes observed for STRUT indicate that it is part of the Reverse Vowel Shift: it is beginning to retract and raise to avoid merger with retracting TRAP. During the process of change, TRAP and STRUT overlap temporarily, leading to occasional realisations that sound merged.

5.2 THE FOOT /ʊ/ VOWEL

In varieties of English in southeast England, FOOT has been cited as leading to short front vowel lowering in a push chain (Torgersen & Kerswill 2004). Short front vowel lowering in Irish English is attributed to what Hickey (2014) calls a ‘rotation principle’, where the short front vowels lower because of back vowel raising (THOUGHT and CHOICE), pushing FOOT, KIT and DRESS away, and pulling TRAP along.⁶⁸ These trends highlight the importance of FOOT in short front vowel lowering, in addition to which the fronting of FOOT in SAfE can be considered an important factor in the Reverse Vowel Shift.

As with the TRAP and DRESS vowels, the FOOT data was coded for following environment, as represented in table 5.2. There are no instances of FOOT occurring before /g, m, n, ŋ/, since words such as these do not exist in SAfE. I have observed that words such as *room* and *broom* are variably pronounced with either the GOOSE or the FOOT vowel (/ru:m/ or /rʊm/) in Cape Town, and these are therefore excluded (Mesthrie 2016, personal communication). The FOOT vowel is not a very common vowel in my sample, occurring in stressed position only 690 times.

Environment	n	Example
_0	307	<i>foot, put</i>
_k	372	<i>book, cook</i>
_l	11	<i>full, pull</i>
TOTAL:	690	

Table 5.2: Number of tokens of FOOT by environment.

The formant plot of these three allophones of FOOT is presented in figure 5.4, and from visual inspection, it would appear that there is a substantial difference between FOOT in the free position (light blue), and FOOT when it occurs before /k/ and /l/ (red and yellow respectively). FOOT_0 is considerably fronter than the other allophones. An lmer analysis of F1 and F2 suggests that there are similarities between the allophones not made evident by the scatter plot. Concerning height, there is no significant difference between FOOT_0 and FOOT_l ($p=0.135$), though FOOT_0 is significantly higher than FOOT_k ($p=0.000$). Along the F2 plane, there are more significant differences between these three allophones: FOOT_0 is the frontest vowel, significantly fronter than FOOT_k ($p=0.015$), though not significantly fronter than FOOT_l ($p=0.921$). These results are problematic, given the clear distinction between

⁶⁸ LOT and STRUT are unaffected in Irish English; and CHOICE is not included in this study.

FOOT_0 and FOOT_1 in figure 5.4. Given that there are only 11 FOOT_1 tokens, this is likely skewing the results of the lmer model.

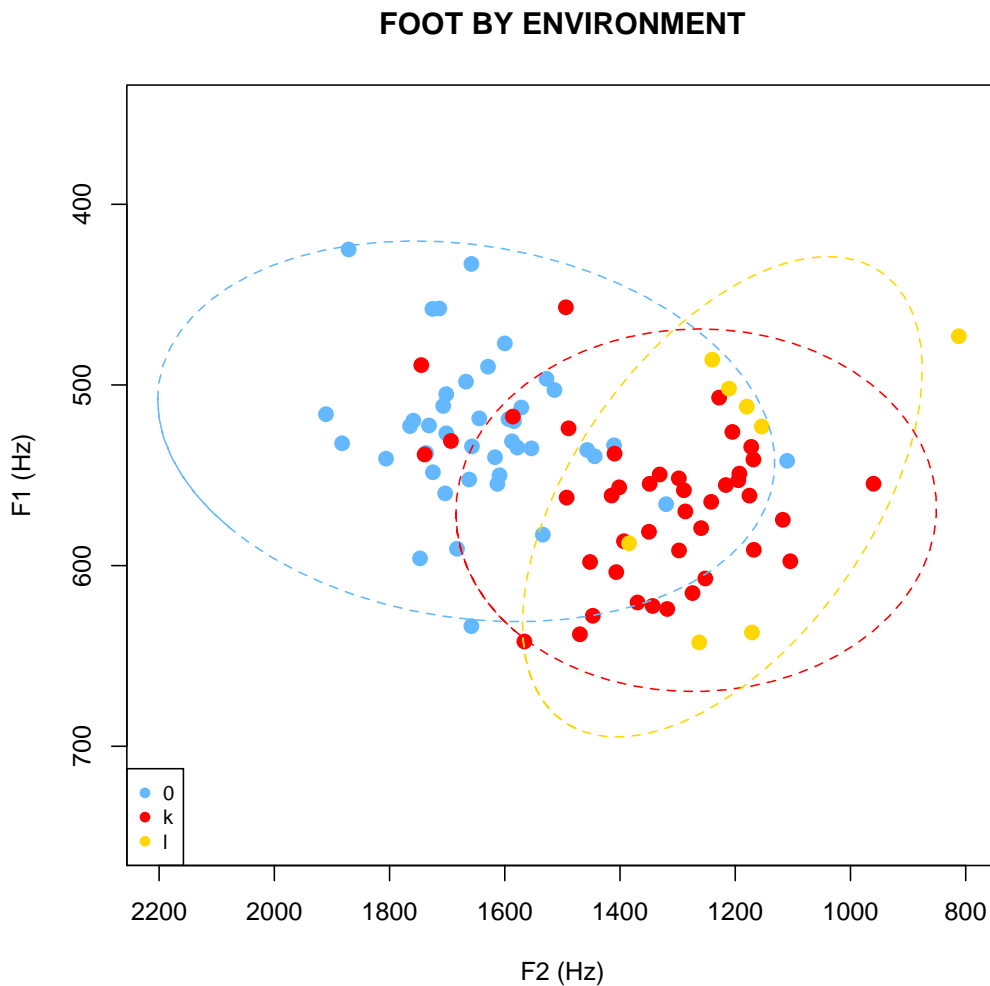


Figure 5.4: Standard deviation (x2) ellipses of the subclasses of the FOOT vowel per environment.

A conditional inference tree analysis for both F1 and F2 return the same results irrespective of whether or not FOOT_1 is included. Figure 5.5 is the ctree analysis for the F1 of FOOT, as predicted by ‘gender’, ‘environment’ and ‘date of birth’. The environment in which the vowel occurs most significantly predicts the height of FOOT. When it occurs before /k/ or /l/ (node 2), it is significantly lower than when it occurs in the free environment (node 3). The realisation of FOOT_0 along the F1 plane is significantly predicted by age: speakers born in or before 1991 have lower realisations of FOOT_0 than speakers born after 1991. Furthermore, the speakers born after 1991 have gendered realisations of FOOT in this environment. The women have lower realisations of the vowel than the men (compare nodes 6 and 7).

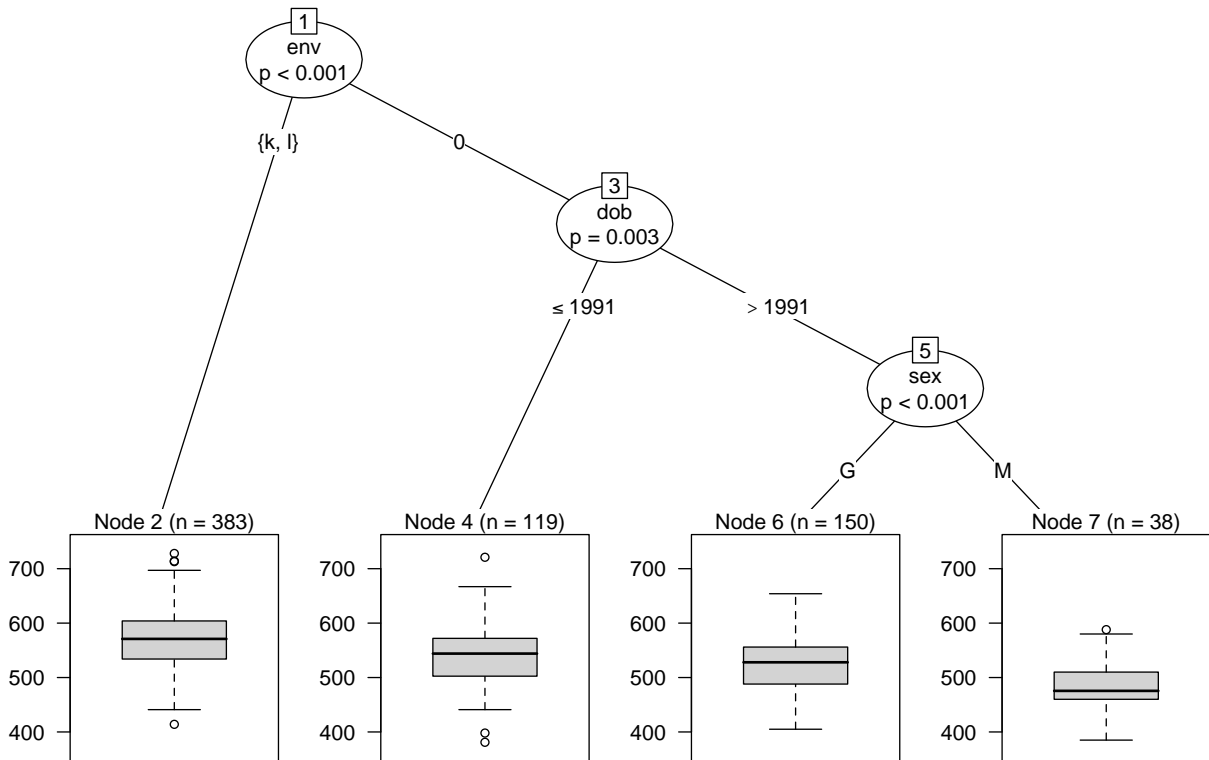


Figure 5.5: Conditional inference tree analysis for FOOT: F1 for ‘gender’, ‘environment’ and ‘date of birth’.

Though the distinction between speakers born in or before 1991 and those born thereafter is very narrow and more than likely not socially significant, the analysis of F1 indicates on a general level that FOOT₀ is raising. Men born after 1991 have higher FOOT₀ than women. In light of Bekker’s (2009) findings of lowered FOOT in formal female speech, and given the large overlap of the box plots in nodes 4 and 6, it seems likely that the men in node 7, who have much higher realisations, resist the lowering of FOOT₀ for some reason. On the other hand, the high realisations in node 7 (which is the major cause of the binary splitting of node 5, and arguably node 3) could also be representative of data skewed by the small number of tokens. The slightly lower realisations in node 2 are expected, since /k/ and /l/ are known as lowering environments.

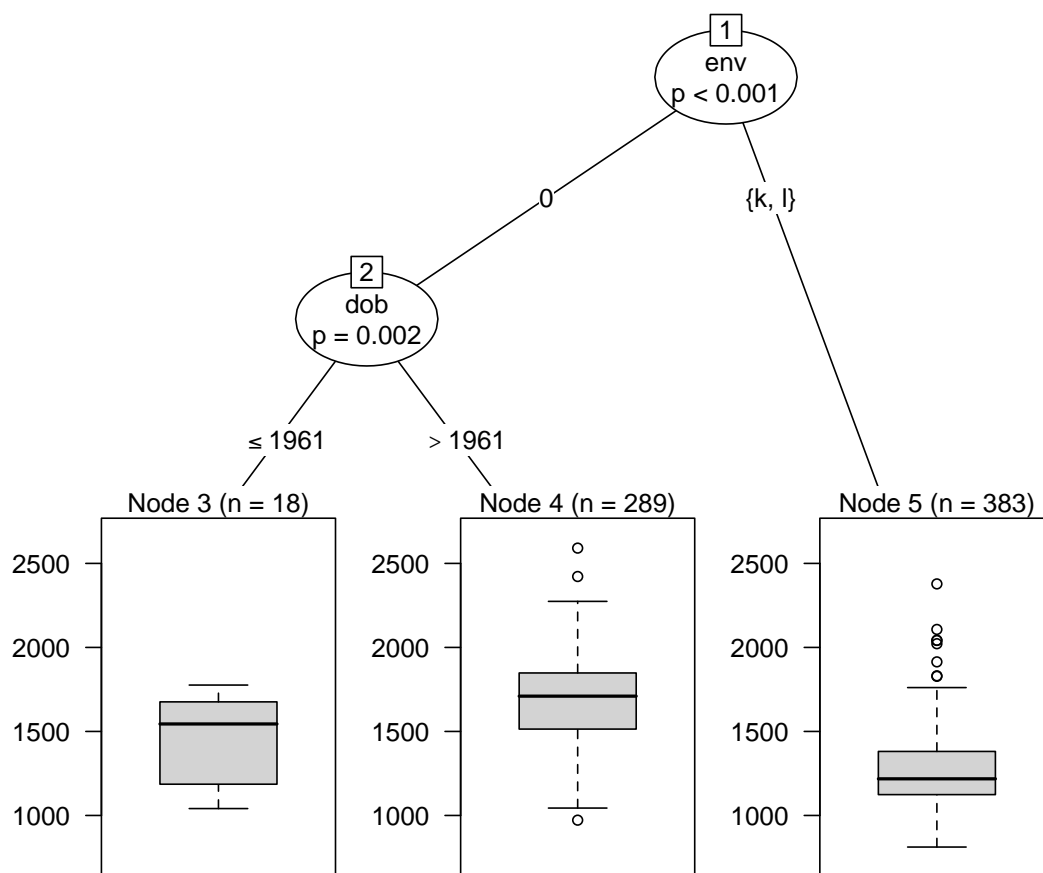


Figure 5.6: Conditional inference tree analysis for FOOT: F2 for 'gender', 'environment' and 'date of birth'.

The situation for the F2 of FOOT is very similar (figure 5.6). FOOT before /k/ and /l/ are the backest realisations of the vowel as expected, uniformly centred around 1300 Hz. When FOOT occurs in the free environment, there is a binary split between the older speakers (node 3) and the younger speakers (node 4). In comparison, the younger speakers front the FOOT vowel significantly, reaching around 2400 Hz where the older speakers only reach about 1700 Hz. Considering the medians, the younger speakers centre around 1700 Hz, whereas the older speakers sit just above 1500 Hz (though the majority are below this median level). The four older speakers, however, only produced 18 FOOT_0 tokens between them, which certainly skews the data. The important pattern here, however, is the consistency with which the younger speakers front FOOT_0. At a median of 1700 Hz, it is a centralised vowel, in line with Lass' (2004, 376) descriptions of a fronter FOOT ([ɥ]) for the younger generations.

As mentioned before, Bekker's (2009, 358-9) findings show a large overlap between KIT_1 and FOOT in all environments except before /l/, which is also found by Lass (1990, 2002). This is corroborated by my casual style data. Figure 5.7 shows a high degree of overlap between KIT6 (KIT before /l/ and after /w/) and FOOT_0.

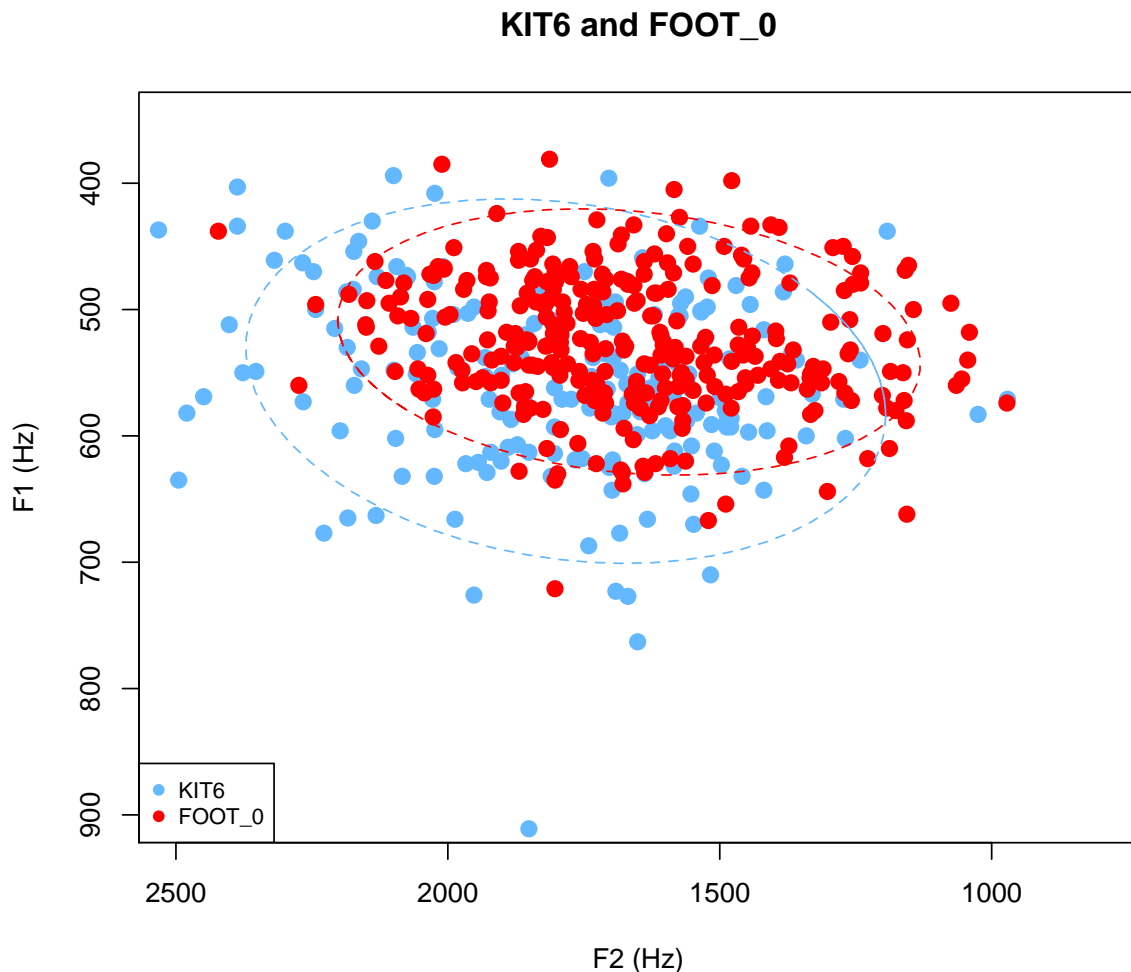


Figure 5.7: Casual style tokens and standard deviation (x2) ellipses for KIT6 and FOOT_0.

A linear mixed effect model for F1 suggests no significant differences between KIT6 and FOOT_0 ($p=0.167$). A ctree analysis, however, finds a significant binary split between the two vowels ($p<0.001$), with KIT6 significantly lower than FOOT_0. Regarding F2, the mixed-effects model returns a significant difference between the vowels ($p<0.000$), with KIT6 400 mean Hz fronter than FOOT_0. This is corroborated by a ctree analysis, which returns a binary split between the vowels, significant at $p<0.001$. (The ctrees are not reproduced here since they are easily explained textually.)

The statistical difference between these vowels in addition to the large overlap between them points to a likely push chain to avoid merger. The fronting of FOOT in the free environment clearly encroaches on the space occupied by KIT. Figure 5.8 is a comparison of FOOT_0 and KIT in all environments, where the imposition of FOOT_0 upon KIT is evident.

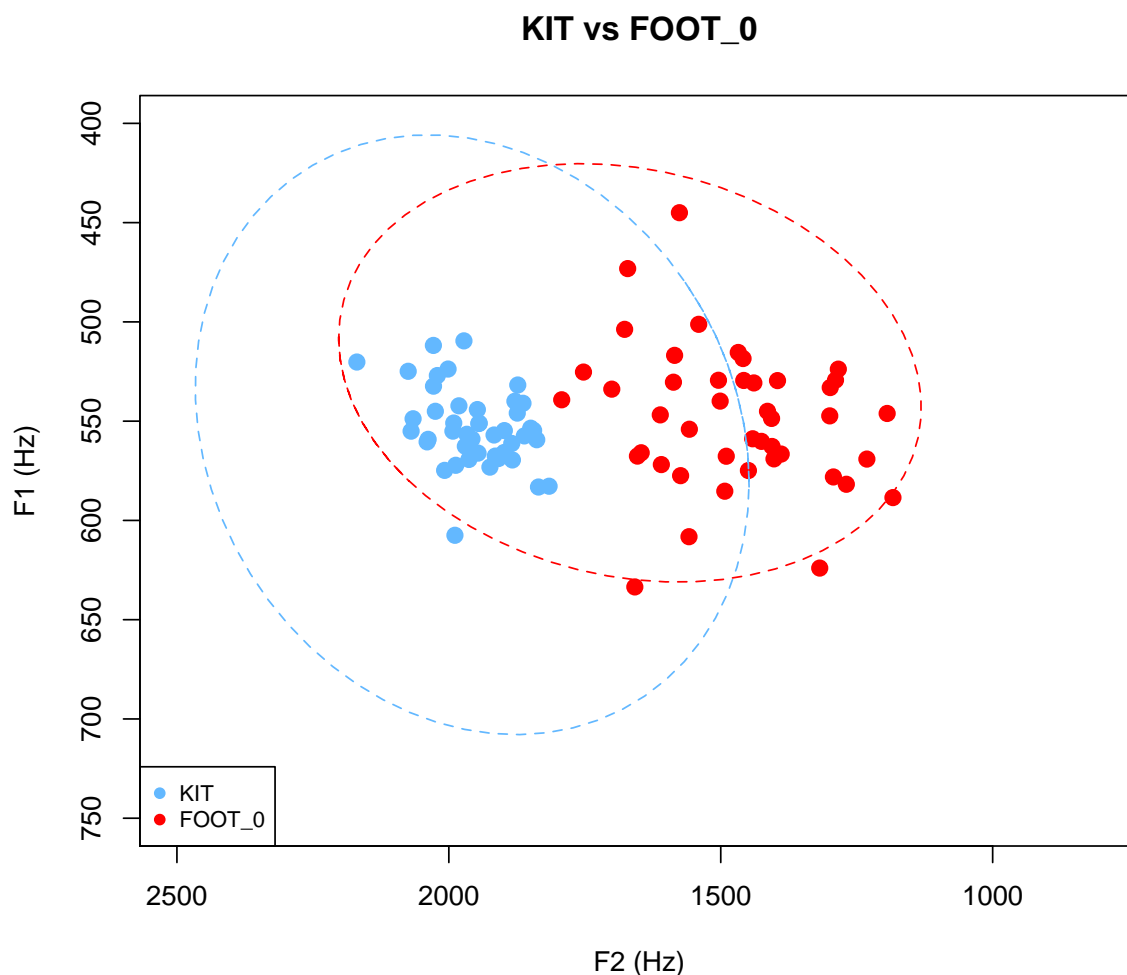


Figure 5.8: Standard deviation (x2) ellipses and mean values of all KIT tokens together with FOOT_0, showing encroachment on KIT's space.

This constitutes motivation for a push chain lowering of the KIT vowel in order to avoid merger with fronting FOOT, a contributing factor to the Reverse Vowel Shift. This possibility is further discussed below, in section 5.5 and in Chapter 6.

5.3 THE LOT /ɒ/ AND THOUGHT /ɔ:/ VOWELS

The characteristics of LOT and THOUGHT, though not the principal concern of this study, are briefly considered here to ascertain if and how they influence the Reverse Vowel Shift. The retraction of TRAP and the subsequent short front vowel lowering in California and Canada have been attributed to Low Back Merger involving LOT and THOUGHT (Chapter 1, section 1.4.3). As this is a brief look at the overall behaviour of these two vowels, the fixed-effect predictors will exclude environment. The focus is instead on gender and age to gain insight into any changes in apparent time that should be studied in detail in the future.

As the ctree in figure 5.9 shows, LOT undergoes lowering over time: younger speakers (the majority of whom are under 30, node 3) have lower LOT realisations than older speakers.

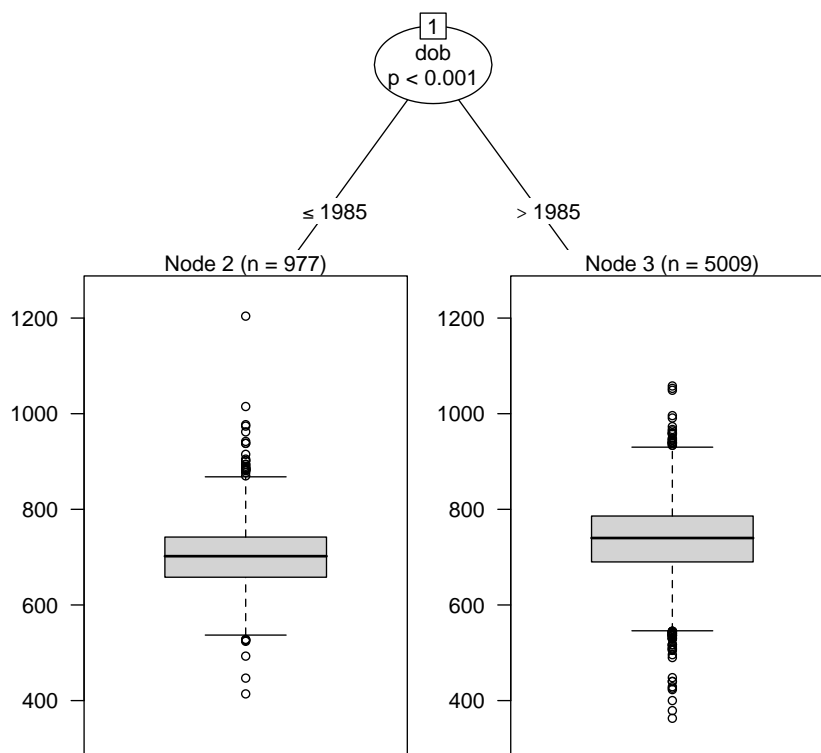


Figure 5.9: Condition inference tree for LOT for F1: ‘gender’ and ‘date of birth’.

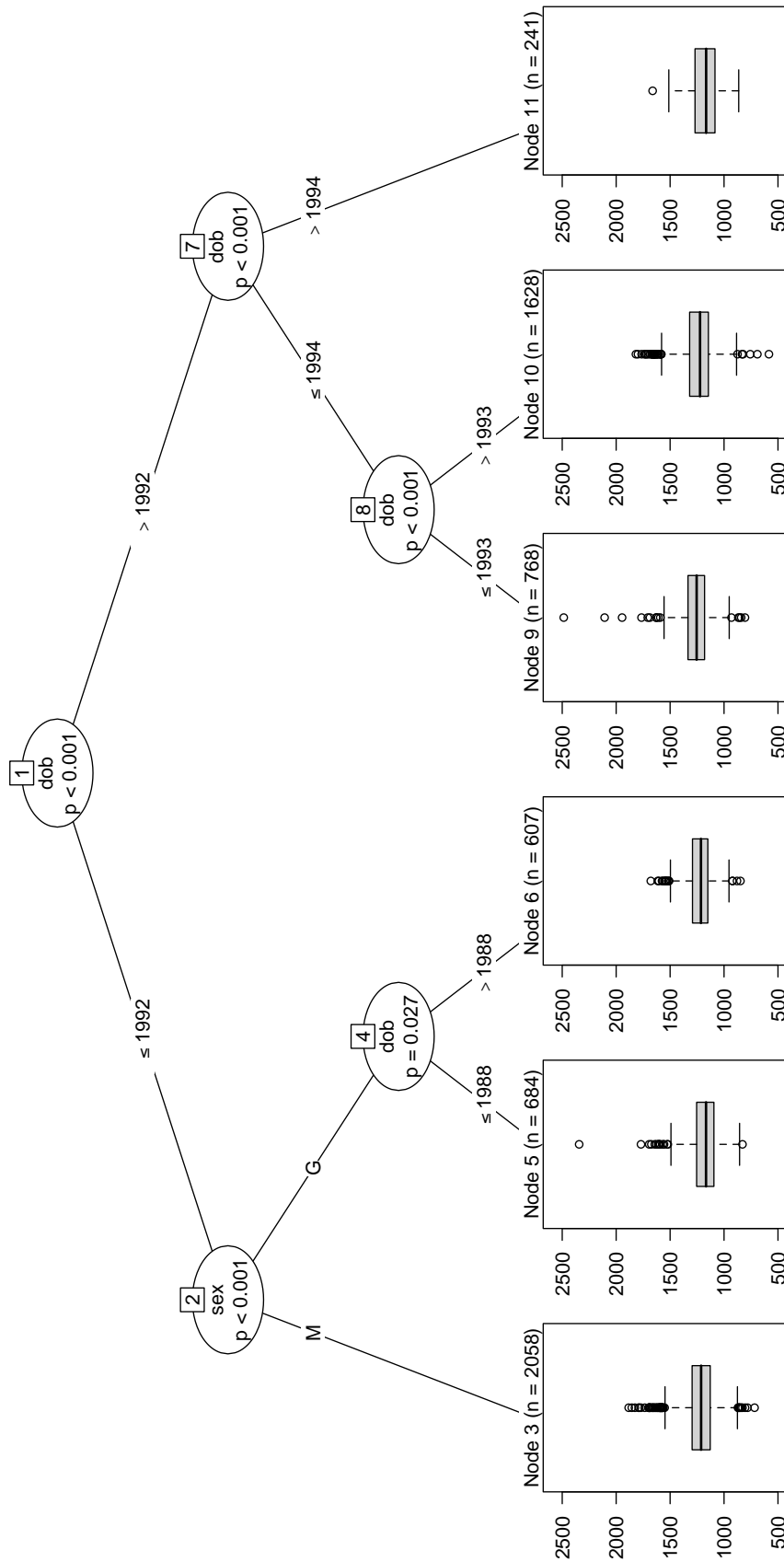


Figure 5.10: Condition inference tree for LOT for F2: 'gender' and 'date of birth'.

The backness of LOT is significantly predicted by age and 'gender'. Though the medians of each of the box plots are very similar, the ctree analysis (figure 5.10) has identified significant differences in predicting factors. Speakers born between 1954 and 1992 (node 2) do not appear very different in their realisations of LOT along the F2 plane. Men between 22 and 60 (node 3) and women between 22 and 58 realise the vowel in a very similar way. The gender distinction is drawn because women between 22 and 25, though having a median similar to the other nodes in this group, generally have fronter LOT. This is evident in the decrease in the spread of tokens represented by the box plot: for the older men and women in this group, LOT realisations go as far back as about 850 Hz, where the younger speakers only reach 1000 Hz.

Speakers born after 1992, i.e. speakers between 19 and 21 years of age, group together. Those born in 1993 are presented in node 9. Node 10 represents speakers born in 1994, and node 11 is CindyH, the only person born in 1995, who mirrors the norms of the older speakers. Speakers between 20 and 21 years of age (1993-1994) have very similar medians to the other groups in the overall sample, though they extend their range of LOT tokens forward: the whiskers of their box plots extend to about 1600 Hz, where the older speakers only reach 1500 Hz.

Given these apparent time generalisations, LOT appears to be lowering and fronting. The fronting of LOT by younger speakers is a feature that has been noted before. As discussed in Chapter 2, LOT is no longer fully back in SAfE, sounding more like [ɫ̩] for younger speakers in Cape Town and KwaZulu Natal (Lass 2004, 376). Bekker (2009, 357-8) also notes the fronted nature of LOT, which he suggests is a feature that has stabilised. The lowering of LOT is perhaps an indication that it may become a fairly low vowel.

Given the fact that LOT appears to lower and front, it provides no cause for TRAP to retract in a pull chain of the sort hypothesised for varieties of American English. Furthermore, it appears unlikely that the lowering and fronting of LOT stimulates the movement of 'old TRAP' downwards and backwards, particularly given the low level of overlap between the two vowels, as presented in figure 4.3 (Chapter 4, section 4.1), repeated below as figure 5.10a. In addition, a chain reaction in vowel movement requires the movement of one vowel in response to the earlier movement of another. The young age of the speakers who apparently

lower and front LOT is the same as those who show the most extreme lowering and retraction of TRAP. This suggests that the movement of LOT is something that occurs in conjunction with or after TRAP lowering. LOT is therefore not a trigger for short front vowel lowering in SAfE as it is in American and Canadian varieties of English.

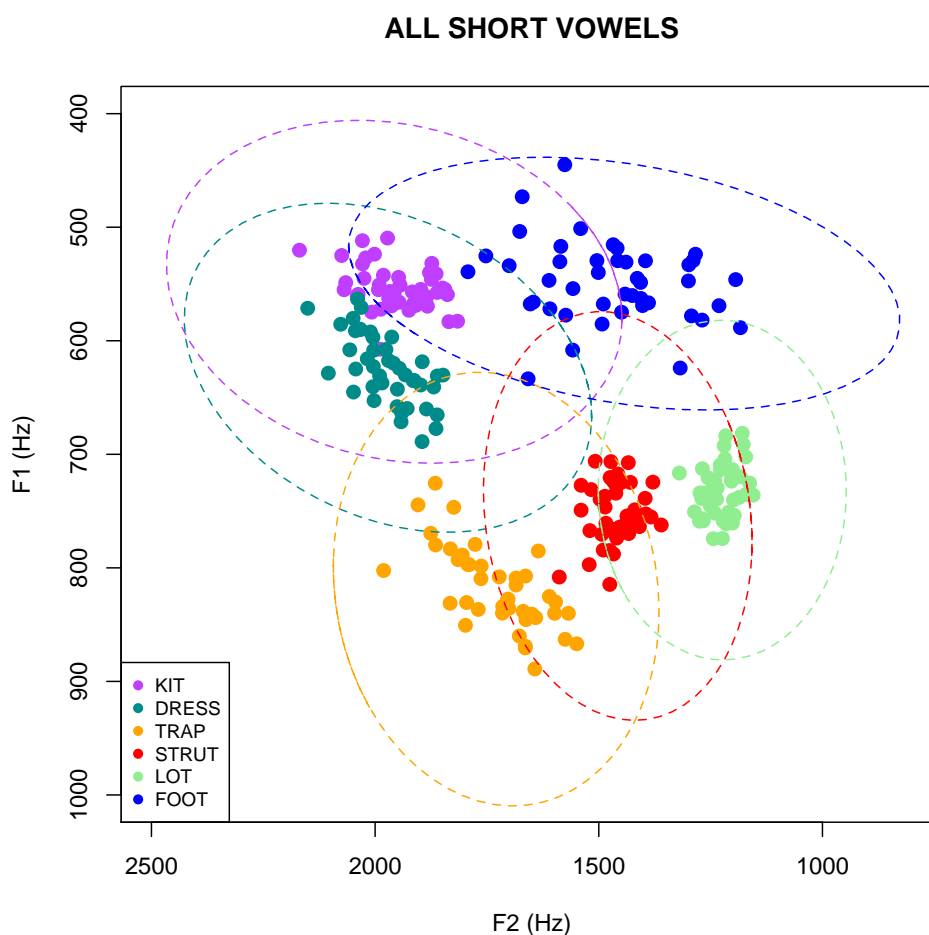


Figure 5.10a: Mean values of all short vowels per speaker with standard deviation (x2) ellipses.

Considering the overlap between LOT and STRUT observed earlier in the chapter, these vowels are likely repositioning, in contrast to TRAP and STRUT as predicted by Bekker (2009). If STRUT raises and retracts, and LOT lowers and fronts, they could trade places, as is visually evident in figure 5.10a. Alternatively, they could come to share vowel space, remaining distinct in terms of rounding: LOT in SAfE is not as rounded as in British English (being weakly rounded), though it is likely rounded enough to maintain a distinction with STRUT.

Though SAfE does not merge LOT and THOUGHT, THOUGHT is considered here to complete the picture. Specifically, THOUGHT as a long vowel may influence TRAP, which has undergone

Lengthening II. THOUGHT is not significantly variable regarding backness, and only shows noteworthy differences for height, presented in figure 5.11. In much the same way as LOT, THOUGHT shows age and gender variation with a degree of overlap between the groups. ‘Gender’ is selected as the fixed-effect that predicts the F1 quality of THOUGHT most significantly ($p < 0.001$). Comparing the men to the women generally, we see that younger women have higher realisations of THOUGHT than men do, evident in higher median values, very close to 600 Hz for men, and closer to 500 Hz for women (especially in nodes 11 and 12).

In addition to the gender distinctions, internal variation is predicted by age. Though the similarities between the men are striking, THOUGHT is generally lower for men born in 1991 (node 5), who have the lowest realisations of THOUGHT overall. Men born after 1991 (i.e. between 1992 and 1994) have generally raised variants in comparison (node 6). Women born between 1954 and 1961 (node 9) have the same median realisation as men born between 1954 and 1990 (node 3). In comparison, younger women have higher realisations, particularly MarilynK in node 11.⁶⁹ THOUGHT thus appears to be raised particularly by young women.

The raised status of THOUGHT in SAfE is well supported in the literature. Descriptions as early as Hopwood (1928) note variation between [ɔ:] and higher [o:] in SAfE. Similarly, Bekker (2009, 312) concludes that THOUGHT has undergone endogenous raising, akin to the patterning observed by Labov’s Pattern III. My data supports the secondary raising in a preliminary sense, particularly in the speech of younger women. Given the fact that the men in my sample generally have lower realisations than the women, it follows that the secondary raising of THOUGHT is still largely in progress, led by women (and could be a chain shift in the making).

⁶⁹ CindyH in node 13 is an exception: even though she is the youngest speaker in the sample; she maintains a lower, older realisation for THOUGHT.

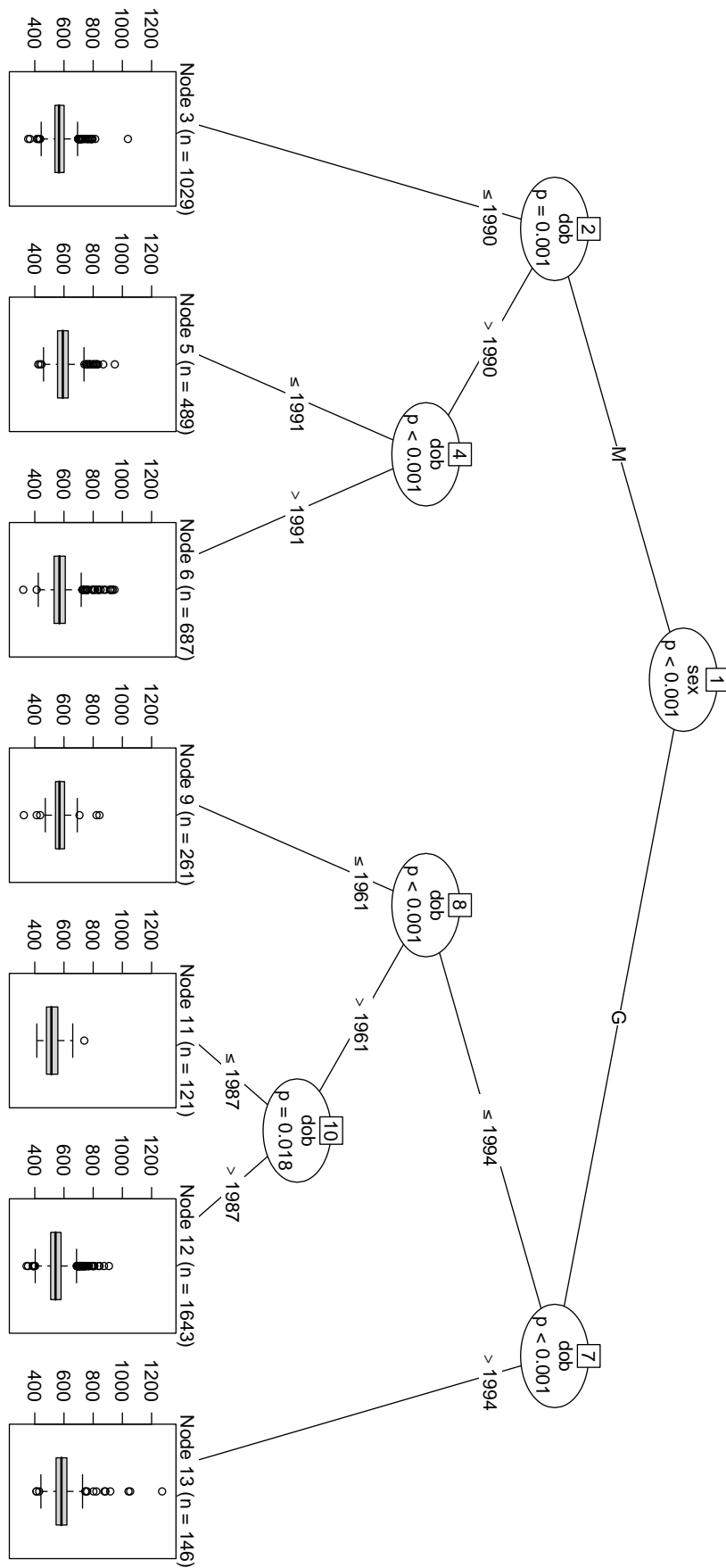


Figure 5.11: Conditional inference tree for THOUGHT for F1: 'gender' and 'date of birth'.

5.4 SUMMATION: ALL SHORT VOWELS AND THOUGHT

Having analysed the impact of STRUT, FOOT, LOT and THOUGHT on the Reverse Vowel Shift, it would appear that the phenomenon involves all the short vowels except LOT. Both FOOT and TRAP have important parts to play in the lowering of the short front vowels. The centralising tendency of FOOT has been noted by numerous early scholars, such as Lanham (1967, 82) and Lass (1990, 277), and the evidence presented in the previous sections supports the fronted nature of FOOT in addition to major overlap with KIT. The encroachment of FOOT on the space occupied by KIT is therefore a plausible cause for the lowering of KIT to avoid merger. It is likely that the fronting of FOOT, an earlier phenomenon than TRAP lowering (noticed in 2000 at the earliest), is responsible for the recent lowering of KIT in a push chain.

The lowering of DRESS is either a pulling effect by TRAP, or a pushing effect by KIT, accommodated because of the space left by lowering TRAP. There is no clarity on whether KIT or DRESS began to lower first, as the lowering of these vowels has not been overtly mentioned in the literature. The analysis presented in this thesis however suggests that the time frame is largely the same, especially given the large overlap between KIT and DRESS. Labov (2010, 140) notes that if two sounds change simultaneously, it is not evidence of a chain shift, but of generalised sound change. Labov (2010, 301) furthermore states that, in phonemic chain shifts, the phoneme only responds to the movement of another phoneme when the latter has vacated its space. If TRAP was the sole instigator for short front vowel lowering, the movement of DRESS downwards would occur first, only *after* which KIT would lower in response. The involvement of FOOT solves this dilemma: KIT lowers because of FOOT fronting, and DRESS lowers because of TRAP lowering and retraction. The Reverse Vowel Shift in SAfE is therefore a combination of a pull and a push chain, involving all the short vowels except LOT.

Important to remember, however, is that these timelines need refinement and further study. Given the uneven weight of my data in terms of age, these apparent time conclusions, though plausible, need further interrogation with a more even sample. A sample with a larger generational span could furthermore provide more conclusive insight into the suggested simultaneous lowering of KIT and DRESS, which based on the current data cannot be ascertained conclusively.

5.5 KIT~DRESS MERGER

The overview of the vowel space in Chapter 4, section 4.1 brought to light the large overlap between the standard deviation ellipses of KIT and DRESS. Though the overlap between KIT and DRESS is likely a temporary feature caused by (what appears to be) simultaneous lowering of KIT and DRESS as part of the Reverse Vowel Shift, the possibility of vowel merger should not be discounted, particularly since these vowels seem to lower in unison. In addition to conditional inference trees and linear mixed-effects models, Pillai Scores, Euclidean Distance measures and Welch's two sample *t*-tests are used in the analysis to follow (these are explained in detail in Chapter 3, section 3.5). The overlap between the two vowel categories is shown in figure 5.12. Though there is considerable overlap between KIT (purple) and DRESS (green), KIT is generally the higher vowel.

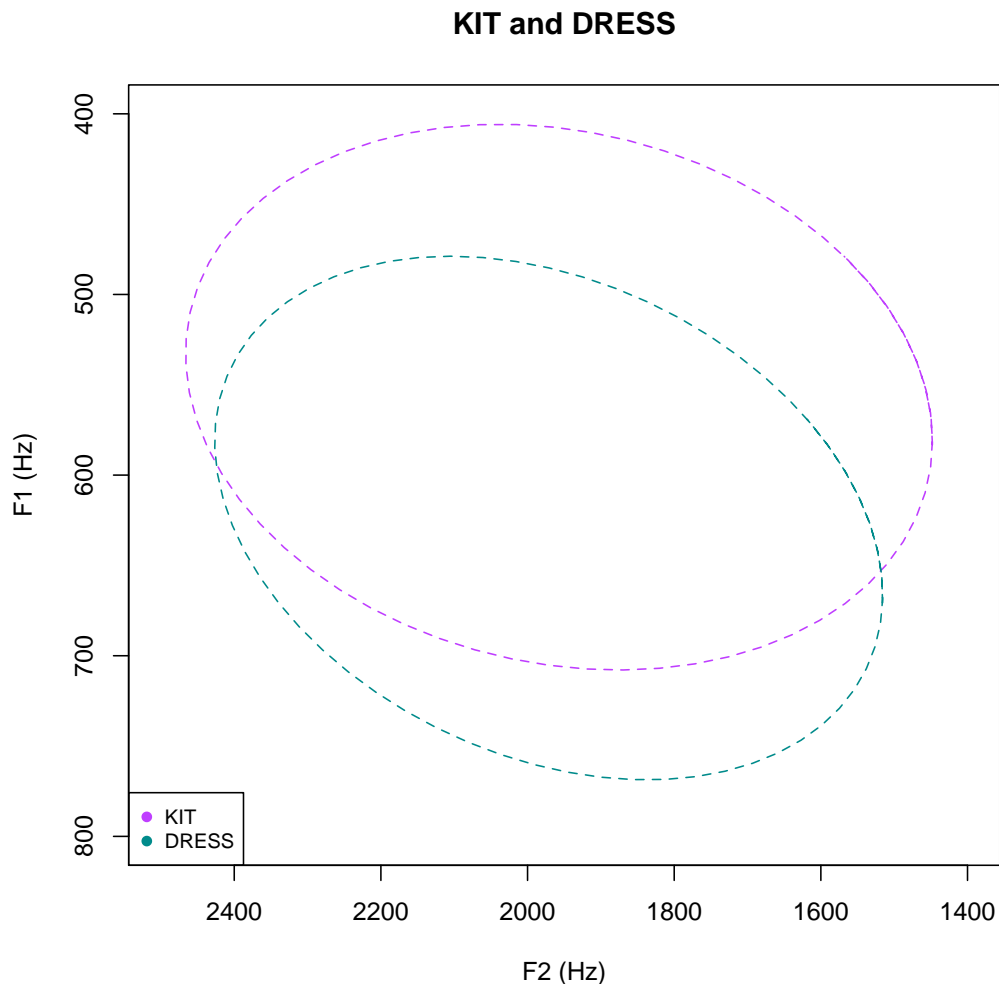


Figure 5.12: Standard deviation (x2) ellipses illustrating the overlap between KIT and DRESS in all environments.

When compared to the results of Bekker's (2009) study (reviewed in Chapter 2, section 2.3.3), it is evident that the positioning of KIT and DRESS in my sample is different. In his data (reproduced in figure 5.13), DRESS and KIT, with their various allophones, are essentially the same height. The standard deviation ellipses in this graph are based on word list tokens and fewer speakers than my graphs, in addition to showing one standard deviation where I show two. Hence, the figure shows smaller distributions when compared to my data. The important illustration provided by this graph is, however, the equal height of KIT and DRESS. Furthermore, DRESS only overlaps KIT2, and then only slightly. The data in my study shows DRESS as not only lower than KIT, but also as overlapping to a large degree across all environments. DRESS has possibly been subject to change since Bekker's study, though it is more likely that the differences are due to style and region: there are no Capetonian speakers in his sample, and he bases his analysis on word list tokens, where mine are causal style. This section therefore explores whether these changes and the large overlap between the categories constitutes vowel merger or temporary overlap related to the Reverse Vowel Shift.

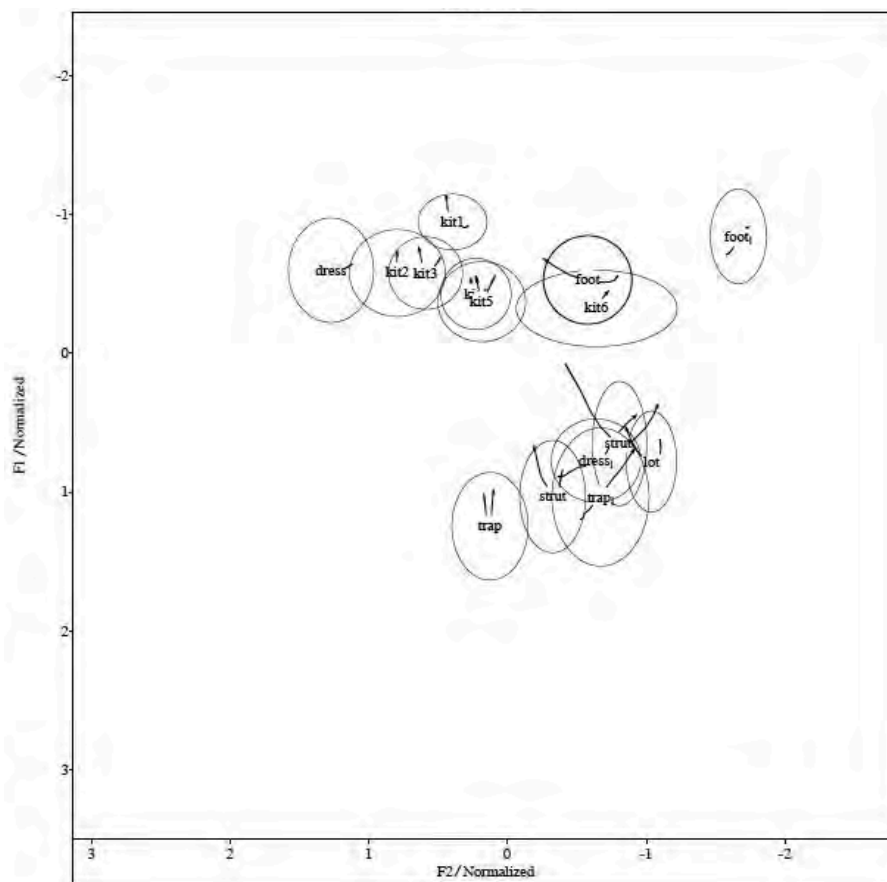


Figure 5.13: Means and standard deviation (x1) of the short vowels of 27 young women in word list style (source: Bekker 2009,197).

In exploring the closeness between KIT and DRESS, the ‘vowel’ category (IH and EH) is added as a fixed-effect predictor in the ctrees analysis.⁷⁰ For F1, KIT and DRESS are pointedly different from the outset. ‘Environment’ emerged as the strongest predicting factor ($p < 0.001$), and the data is separated with all the DRESS environments branching to the left, and all the KIT environments to the right. A conditional inference tree for F2, however, shows that ‘environment’ is the strongest predicting factor. KIT2 is grouped with most of the DRESS environments. DRESS_l and DRESS_m are grouped with KIT0, KIT3, KIT5, and KIT6. Though DRESS_l initially patterns with KIT, it is subsequently identified as different. A mixed-effects regression model also indicates that KIT and DRESS_l are significantly different from one another along F1 and F2 ($p < 0.000$).

In section 4.3, DRESS_m is identified as the first environment where DRESS undergoes retraction. Figure 5.14 shows that it groups significantly with all ‘other’ KIT environments (i.e. excluding KIT2). The exact way in which the vowels overlap is predicted by ‘gender’: men have an overlap between DRESS_m, KIT 3 and KIT5 (node 4), while women have an overlap between KIT0, KIT 3 and DRESS_m (node 6).

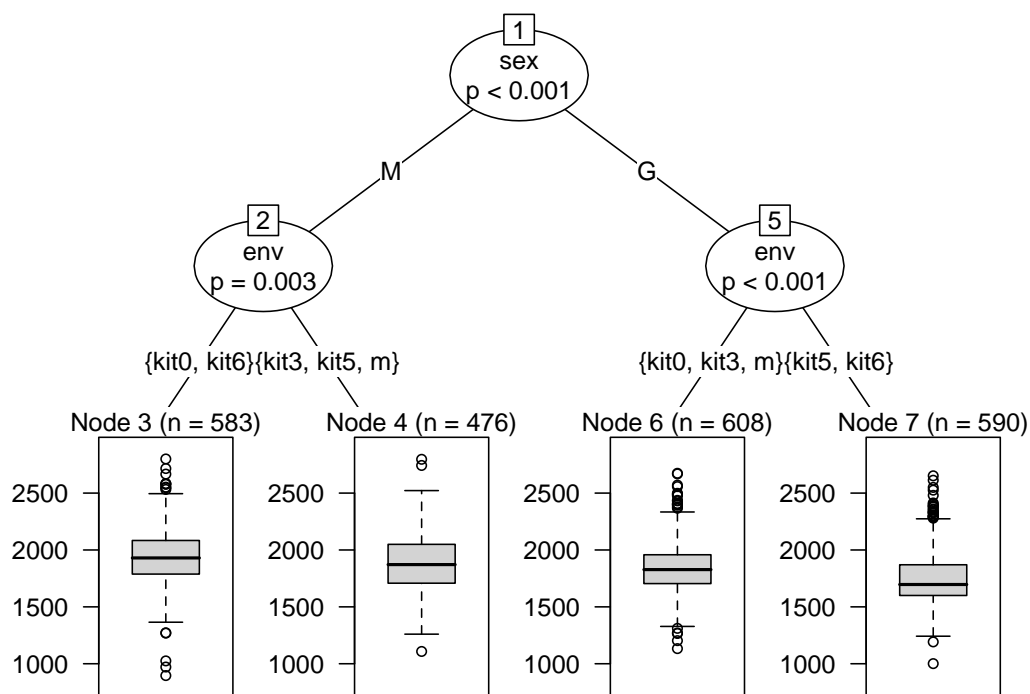


Figure 5.14: Conditional inference tree for retracted KIT and DRESS_m: F2 for ‘gender’, ‘environment’, ‘vowel’ and ‘date of birth’.

⁷⁰ These ctrees are not reproduced here given their size and complexity.

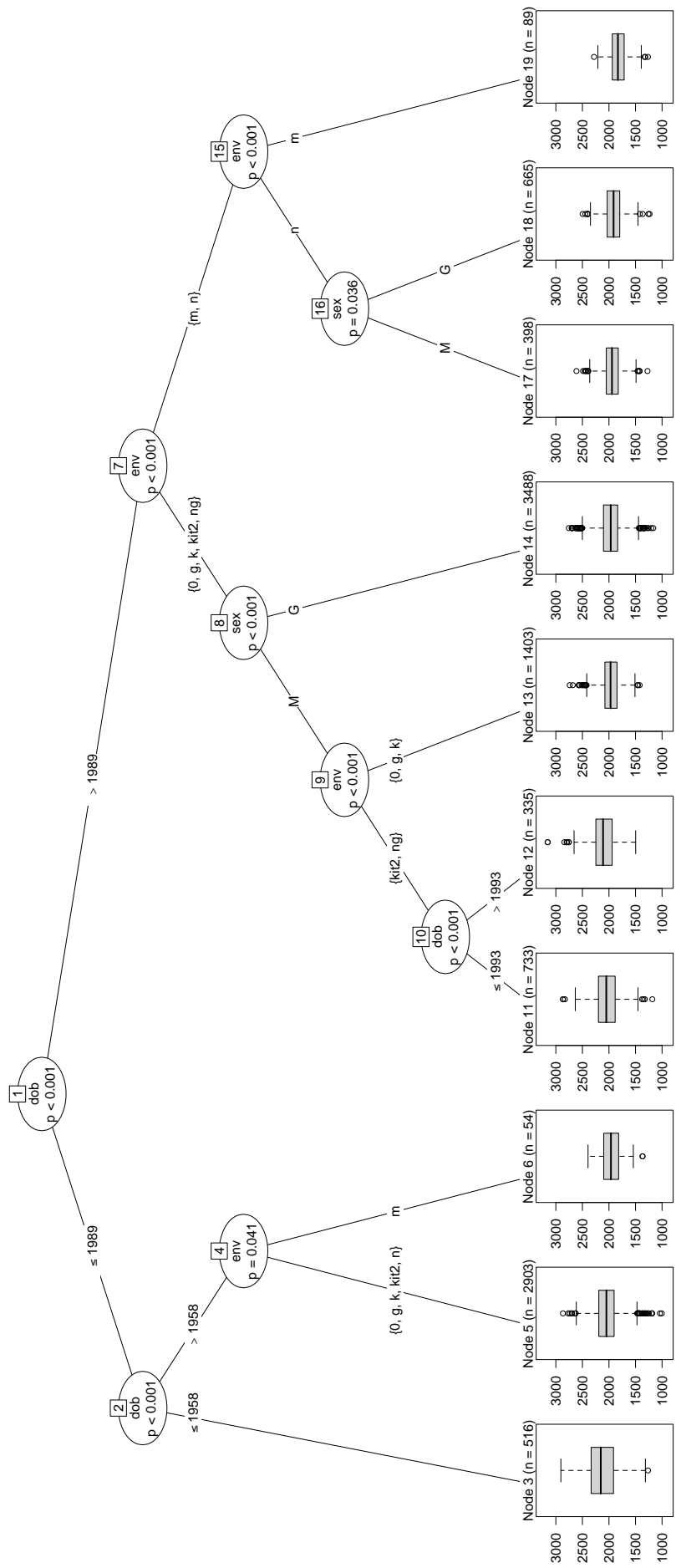


Figure 5.15: Conditional inference tree KIT2 & DRESS (excluding /l/) F2: ‘gender’, ‘environment’ and ‘date of birth’.

The ctree alluded to earlier indicated an overlap between KIT2 and DRESS excluding DRESS_I. The results of the ctree in figure 5.15, showing the analysis for F2, are quite complex, and are thus summarised as follows:

1. KIT2 patterns with DRESS, predicted by age.
2. People born in or before 1989 make no distinction between KIT2 and DRESS, except in the case of DRESS_m (node 6).
3. People born after 1989 have conditioned distinctions in F2 between the vowel classes:
 - a. DRESS_m and DRESS_n do not pattern with KIT2 (node 15).
 - b. The rest of the environments pattern together, and are conditioned by gender.
 - i. Women make no distinction between KIT and DRESS (node 14).
 - ii. Men group KIT2 and DRESS_η together (node 10).

While there are significant similarities between KIT and DRESS for F2 it is of little consequence at this stage since there is no significant overlap indicated for F1, and merger depends on both. Nevertheless, the large degree of overlap between DRESS and KIT necessitates further analysis. Pillai Scores and Euclidean Distance measures, though they do not take environment into account in as much detail as the ctrees, consider F1 and F2 simultaneously, and provide a measure of distance between the vowels per speaker. This illuminates how the individual treats these vowel classes and provides an integrated analysis of F1 and F2. The data is separated into three sets, based on the environmental indications obtained from the ctrees:

Set 1: KIT and DRESS.

Set 2: KIT and DRESS excluding /I/.

Set 3: KIT2 and DRESS excluding /I/.⁷¹

Of the 44 speakers in the sample, four speakers have very low Pillai Scores (under 0.1) for all three sets, as is shown in table 5.3. The vowels in Set 3 are closer to one another than the vowels in the other sets (for all except LexiK). This is particularly true for MarilynK, who has fairly high Euclidean Distances between the vowels in Set 1 and Set 2 and much lower values for Set 3 (particularly the Pillai Score). In fact, the *p*-value for Set 3 in MarilynK's results shows that the Pillai distance between the vowels is not significant. The other speakers maintain a significant distinction between the vowels in each set.

⁷¹ DRESS_m is not investigated as a separate set with 'other' KIT due to the small number of tokens of DRESS_m: n = 149.

Speaker	Test	Set 1	Set 2	Set 3
JaneK	Pillai	0.072	0.074	0.044
	p	<0.000	<0.000	<0.000
	Euclidean	53	80	43
LexiK	Pillai	0.090	0.087	0.089
	p	<0.000	<0.000	<0.000
	Euclidean	58	60	53
MarilynK	Pillai	0.050	0.054	0.003
	p	<0.000	<0.000	0.61
	Euclidean	96	112	20
MilesG	Pillai	0.074	0.060	0.030
	p	<0.000	<0.000	0.012
	Euclidean	66	56	61

Table 5.3: Distance measures between close KIT and DRESS.

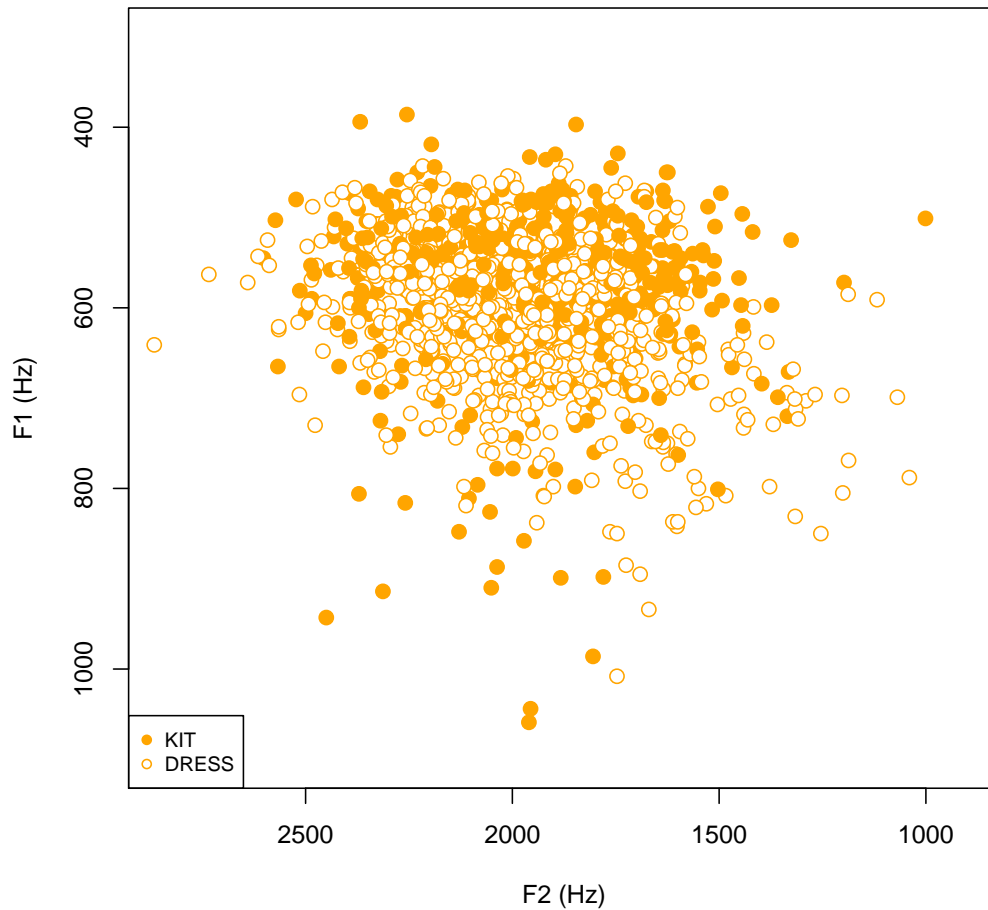


Figure 5.16: Very close KIT and DRESS tokens of 4 speakers with Pillai Scores under 0.1.

The very low Pillai Scores and the small Euclidean Distance in hertz between the vowels, mostly well under 100, the threshold for merger as posited by Evanini (2009), show that the distinction between these vowels for these four speakers, though significant, is very small if not merged, as the formant plot in figure 5.16 shows. To further test the closeness between KIT and DRESS for these 4 speakers, table 5.4 provides the results for Welch Two Sample *t*-tests, comparing the F1 and F2 of KIT and DRESS separately. Only non-significant *p*-values are provided.

Speaker	Formant	Set 1 <i>p</i> -value	Set 2 <i>p</i> -value	Set 3 <i>p</i> -value
JaneK	F1			
	F2			0.071
LexiK	F1			
	F2	0.233		0.536
MarilynK	F1			0.312
	F2		0.224	0.60
MilesG	F1			
	F2	0.064	0.936	

Table 5.4: Comparison of KIT and DRESS: results of Welch's *t*-tests.

A similar pattern is observed here when compared to the Pillai Scores and Euclidean Distance measures provided in table 5.2: the vowels in Set 3 are the closest together (given the increase in non-significant differences for F1 and F2). MarilynK, in fact, shows no significant difference between KIT and DRESS in Set 3. In addition, she has very low Pillai Scores and Euclidean Distances less than 100 Hz (table 5.3). Given these results, she displays conditioned merger of KIT and DRESS.

On the opposite side of the scale, two speakers show greater difference between the three sets, with Pillai Scores of between 0.5 and 0.7, and Euclidean Distances over 200 Hz, presented in table 5.5. As the format plot (figure 5.17) shows, these speakers have largely separate realisations for KIT and DRESS. In addition, the degree of distance between KIT and DRESS is not phonologically conditioned. This group has the least amount of similarity between KIT2 and DRESS excluding /ɪ/ (Set 3), with highly significant differences between the vowels in each set.

Speaker	Test	Set 1	Set 2	Set 3
AlanL	Pillai	0.528	0.532	0.544
	p	<0.000	<0.000	<0.000
	Euclidean	218Hz	207Hz	289Hz
LarryR	Pillai	0.602	0.588	0.620
	p	<0.000	<0.000	<0.000
	Euclidean	340Hz	330Hz	378Hz

Table 5.5: Distance measures between distinct KIT and DRESS.

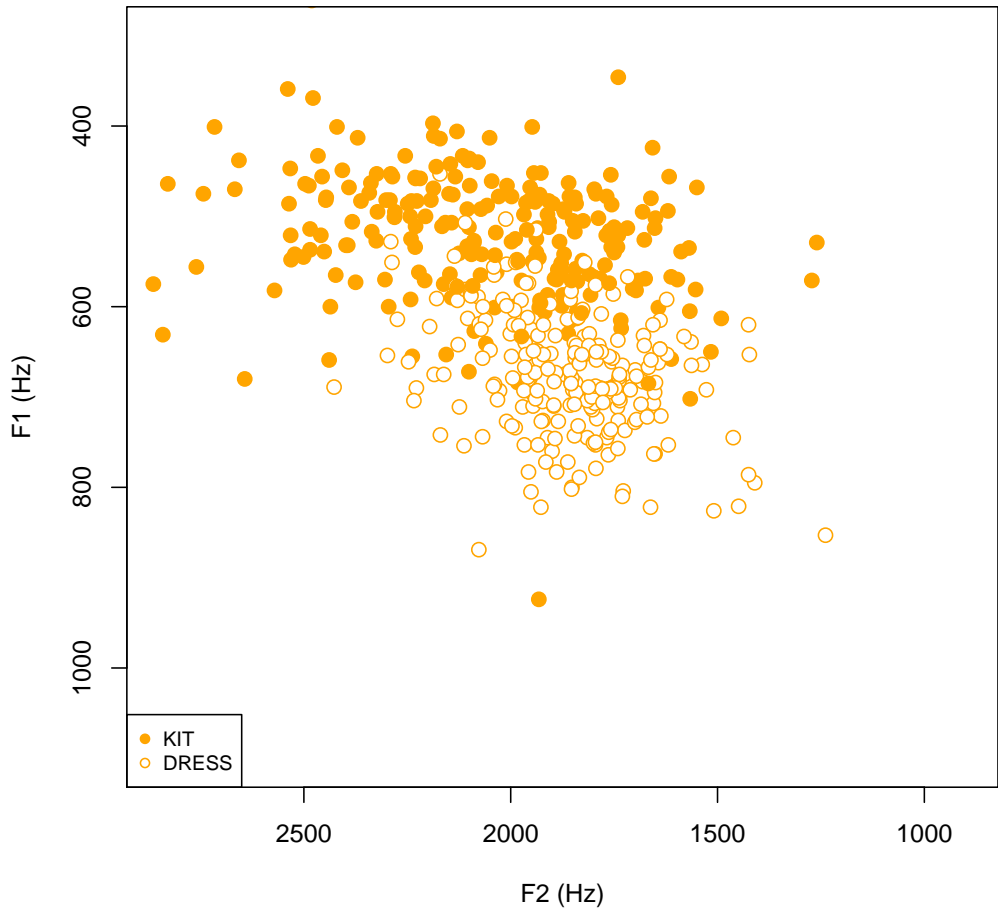


Figure 5.17: Distinct KIT and DRESS of 2 speakers with Pillai Scores over 0.5.

Between these two extremes lie the remaining 38 speakers, who return Pillai Scores ranging from 0.1 to 0.5. These speakers have close KIT and DRESS realisations as is seen in figure 5.18. In general, Pillai Scores are highest for the set which includes all environments (Set 1), and lowest for the set only including KIT2 and DRESS excluding 1 (Set 3). The set including all KIT environments and DRESS excluding 1 (Set 2) falls in between.

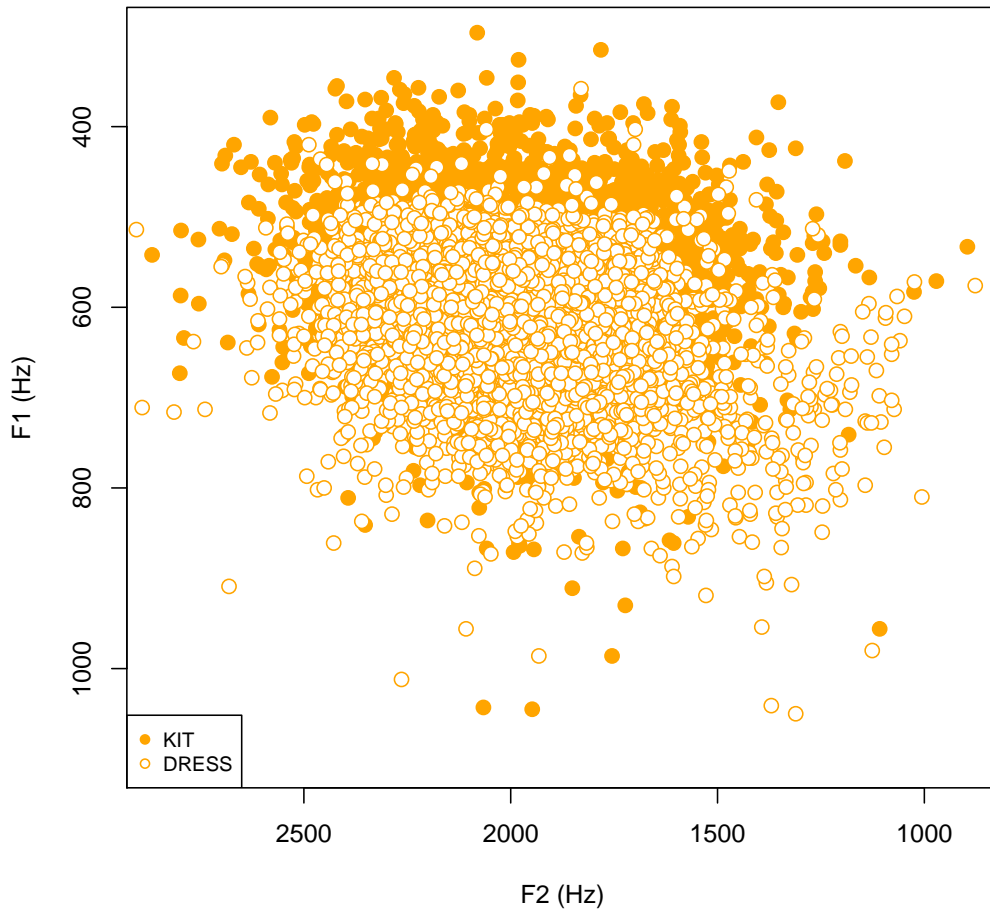


Figure 5.18: Close KIT and DRESS for 38 speakers with Pillai Scores between 0.1 and 0.5.

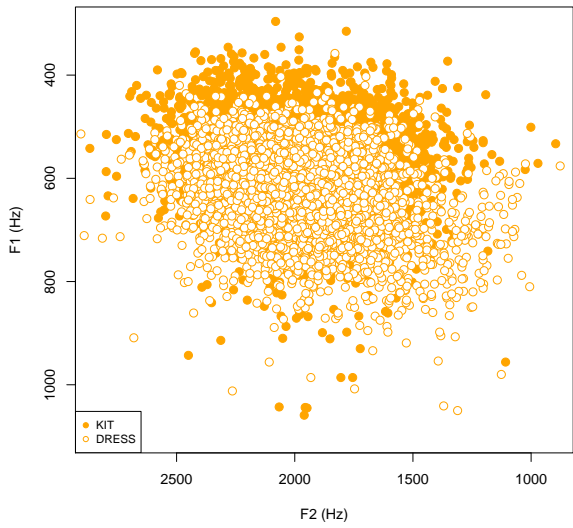
Of these 38, three speakers produce the vowels in Set 3 so close together that they are not significantly different, presented in table 5.6. Though these three speakers have Pillai Scores over 0.1 for sets 1 and 2, they fall to well below 1 for Set 3 in addition to showing no significant difference. Furthermore, the Euclidean Distance between KIT and DRESS in Set 3 for these speakers is well below 100 Hz, coupled with non-significant p -values returned by a Welch t -test.

Speaker	Mean	KIT	DRESS	Euclidean Distance	p Welch	Pillai Score	p
BobbyC	F1	578	585	61 Hz	0.24	0.030	0.123
	F2	1960	2018		0.19		
BrendaF	F1	542	566	43 Hz	0.16	0.036	0.314
	F2	2120	2156		0.62		
KayC	F1	562	577	39 Hz	0.07	0.025	0.07
	F2	2034	2070		0.25		

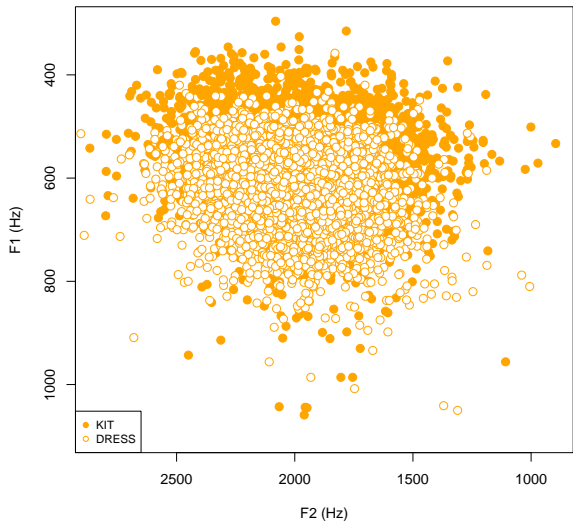
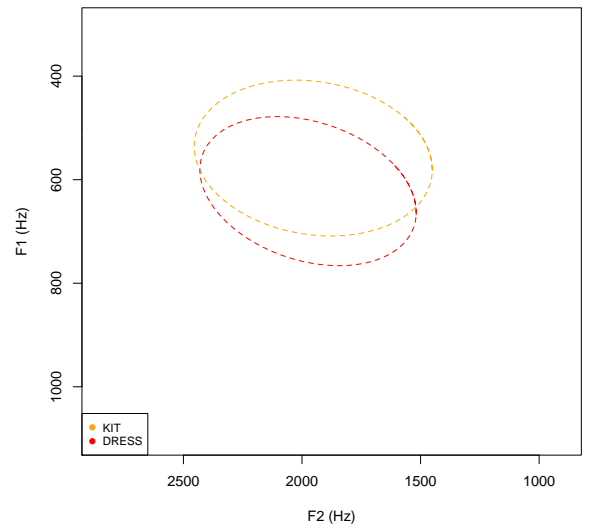
Table 5.6: Distance measure for KIT and DRESS Set 3 for 3 speakers.

In addition to MarilynK, therefore, these speakers merge KIT2 and DRESS excluding /I/ (Set 3). The increased closeness between KIT2 and DRESS excluding /I/ is supported by the fact that, of the 44 speakers, only 12 have a distance of more than 100 Hz between the vowels in Set 3 (compared to 22 for Set 1 and 25 for Set 2). This, combined with the conditioned distance between KIT and DRESS (presented in figure 5.19) and the low Pillai Scores for many of the speakers strongly suggests that these two vowels are very close to one another and merge for some speakers, conditioned to Set 3.

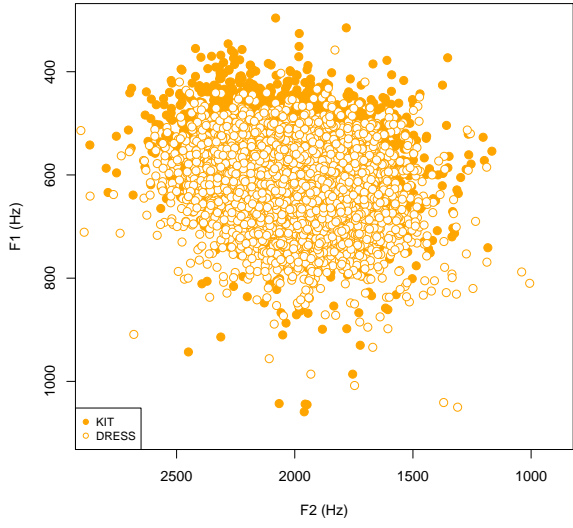
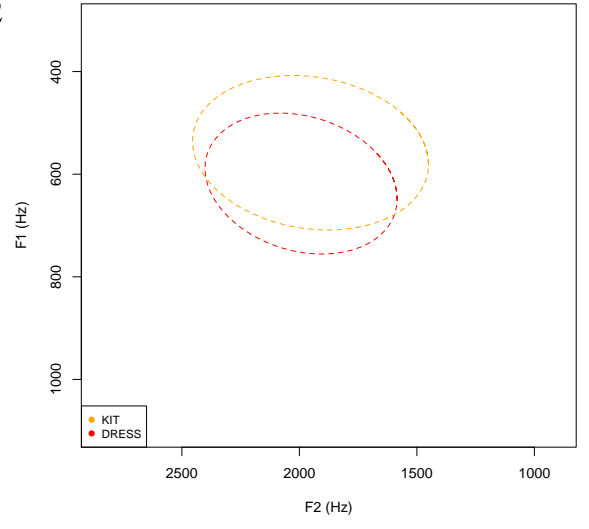
Given the fact that the closeness between these two vowels was not anticipated, no minimal pairs testing was carried out. The proposed conditioned merger is therefore preliminary at most. In addition, only four of the 44 speakers indicate the conditioned merger. This either indicates a change in progress, or an overlap of vowel categories as a result of the Reverse Vowel Shift. There is no clear social pattern to the ‘merger’: the youngest speaker to participate is MarilynK (born in 1987), who has been a continual outlier when compared to her peer group. Two older women indicate this merger KayC (1961) and BrendaF (1957), and one man, BobbyC, born in 1983. Given the fact that no speakers born between 1987 and 1995 participate, it follows that this is not a change in progress. It is therefore more likely that this is a temporary feature, perhaps associated with the Reverse Vowel Shift.



SET 1



SET 2



SET 3

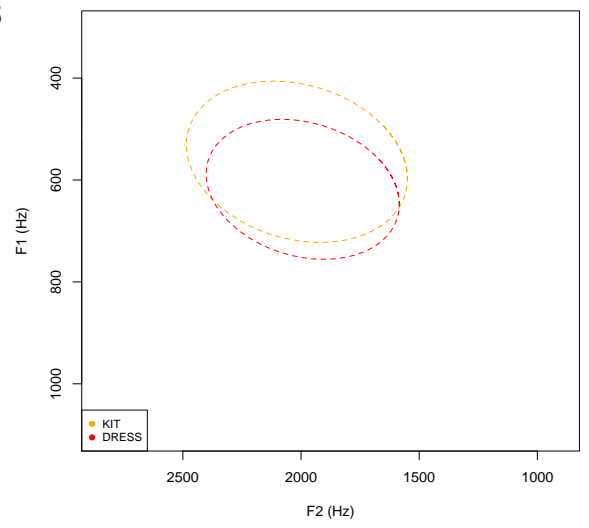


Figure 5.19: Conditioned distance between KIT and DRESS: all tokens (left) and standard deviation (x2) ellipses showing overlap (right).

5.6 TRAP~STRUT MERGER

The merger of TRAP and STRUT was identified as a possible change in progress at the beginning of this study. This section presents the results of the detailed analysis into the possibility of merger. The analysis starts with the casual style data, and is supplemented by minimal pairs testing. The tools used to analyse closeness between the vowel categories remain Pillai Scores and Euclidian Distance measures. In addition, Welch's two sample t -tests are applied to the tokens to further ascertain the significance of distance.

Conditional inference trees indicate that there are no significant similarities between TRAP and STRUT for either F1 or F2. An exception is when STRUT occurs before /g/, which is indistinguishable from TRAP. However, there are only seven instances of STRUT_g: *dug*, *bug/s*, *hug*, *juggling*, *mug* and *ugly*. Therefore, no concrete claims can be made about this, and I exclude these as a result. Mixed-effects regression models also show that TRAP and STRUT are significantly different from one another ($p < 2e-16$ for both F1 and F2).

The Pillai Scores and Euclidean Distance measures (together with their significances) for speakers with the smallest distance between TRAP and STRUT are presented in table 5.7. Of the 44 speakers, 21 return a Pillai Score under 0.5.⁷² Though the Pillai Score is fairly low for these speakers (though not lower than 0.2), all p -values show that these speakers keep the vowels significantly distinct. Moreover, the Euclidean Distance between the vowels is well over 100 Hz for each speaker. As evident in the scatterplot (figure 5.20) and the standard deviation ellipses (figure 5.21), these vowels overlap, though they do not overlap in a way suggestive of merger.

⁷² Results for all speakers are in Appendix C.

Speaker	Mean	TRAP	STRUT	Euclidean Distance	<i>t</i> -score	<i>p</i> Welch	Pillai Score	<i>p</i>
BobbyC	F1	789	808	217	-1.227	0.223	0.442	0.000
	F2	1805	1588		7.025	0.000		
BradP	F1	834	774	264	5.695	0.000	0.479	0.000
	F2	1715	1460		11.998	0.000		
ChloëV	F1	830	746	139	4.518	0.000	0.335	0.000
	F2	1597	1486		5.383	0.000		
EricaV	F1	867	725	174	7.451	0.000	0.382	0.000
	F2	1549	1450		3.60	0.001		
GaryL	F1	870	753	294	7.622	0.000	0.471	0.000
	F2	1664	1395		11.302	0.000		
HayleyS	F1	827	764	257	5.090	0.000	0.416	0.000
	F2	1703	1454		9.879	0.000		
IlseP	F1	840	737	131	9.082	0.000	0.278	0.000
	F2	1567	1486		3.801	0.000		
JaneK	F1	838	731	186	11.139	0.000	0.453	0.000
	F2	1669	1517		11.702	0.000		
JennyL	F1	825	723	207	6.482	0.000	0.466	0.000
	F2	1611	1429		8.690	0.000		
KaleyW	F1	815	728	241	4.806	0.000	0.431	0.000
	F2	1685	1460		9.173	0.000		
KaraL	F1	809	749	157	4.071	0.000	0.302	0.000
	F2	1685	1540		7.036	0.000		
LarryR	F1	807	706	186	5.233	0.000	0.359	0.000
	F2	1664	1508		3.583	0.001		
LauraE	F1	899	785	197	5.138	0.000	0.452	0.000
	F2	1643	1475		8.723	0.000		
LeeL	F1	841	770	181	5.664	0.000	0.456	0.000
	F2	1660	1493		9.778	0.000		
LexiK	F1	860	766	233	4.467	0.000	0.447	0.000
	F2	1678	1464		6.783	0.000		
MindyG	F1	846	797	149	3.556	0.001	0.321	0.000
	F2	1663	1521		8.298	0.000		
PerryF	F1	747	727	285	0.704	0.490	0.424	0.000
	F2	1824	1540		5.624	0.000		
RichS	F1	797	717	342	7.292	0.000	0.447	0.000
	F2	1790	1457		10.891	0.000		
RuthW	F1	840	749	201	5.182	0.000	0.465	0.000
	F2	1599	1420		8.758	0.000		
SarahN	F1	828	724	259	5.572	0.000	0.407	0.000
	F2	1704	1467		6.612	0.000		
TodK	F1	808	770	260	3.90	0.000	0.491	0.000
	F2	1722	1465		14.349	0.000		

Table 5.7: Distance and significance measures for the closest tokens of casual style TRAP and STRUT.

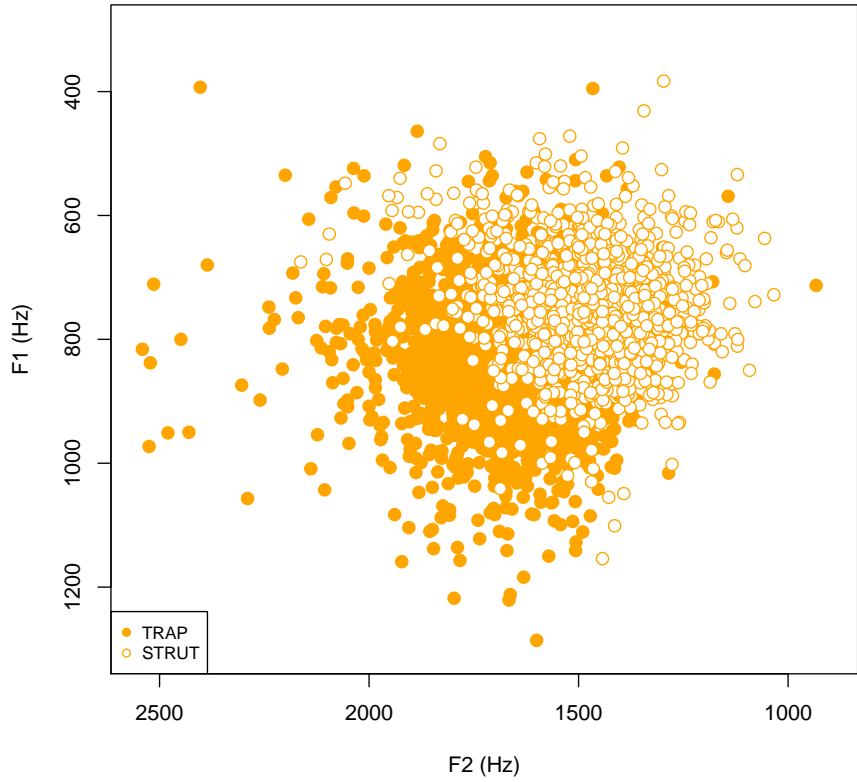


Figure 5.20: All casual style tokens of TRAP and STRUT with Pillai Scores between 0.2 and 0.5.

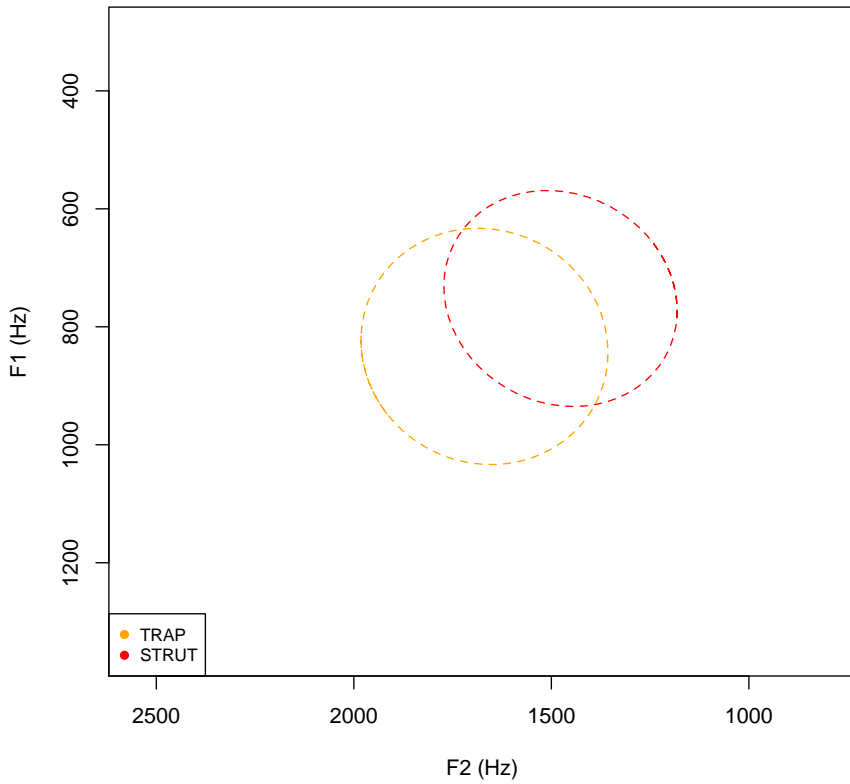


Figure 5.21: Standard deviation (x2) ellipses of all casual style tokens of TRAP and STRUT with Pillai Scores between 0.2 and 0.5.

The remaining 23 speakers return Pillai Scores between 0.5 and 0.8 (figures 5.22 and 5.23), and thus maintain a significant distinction between TRAP and STRUT despite the overlap. There is no pattern regarding social correlates of the distance measures. Those with Pillai Scores below 0.5 were born between 1954 and 1994, and those with Pillai Scores above 0.5 were born between 1956 and 1995. Both groups also have members of both genders. Given the insignificance of the overlap between these vowel classes, and the fact that TRAP and STRUT are both retracting in their own right, there is no merger in progress: TRAP pushes STRUT away in a chain shift process leading to temporary, though significantly different, overlap.

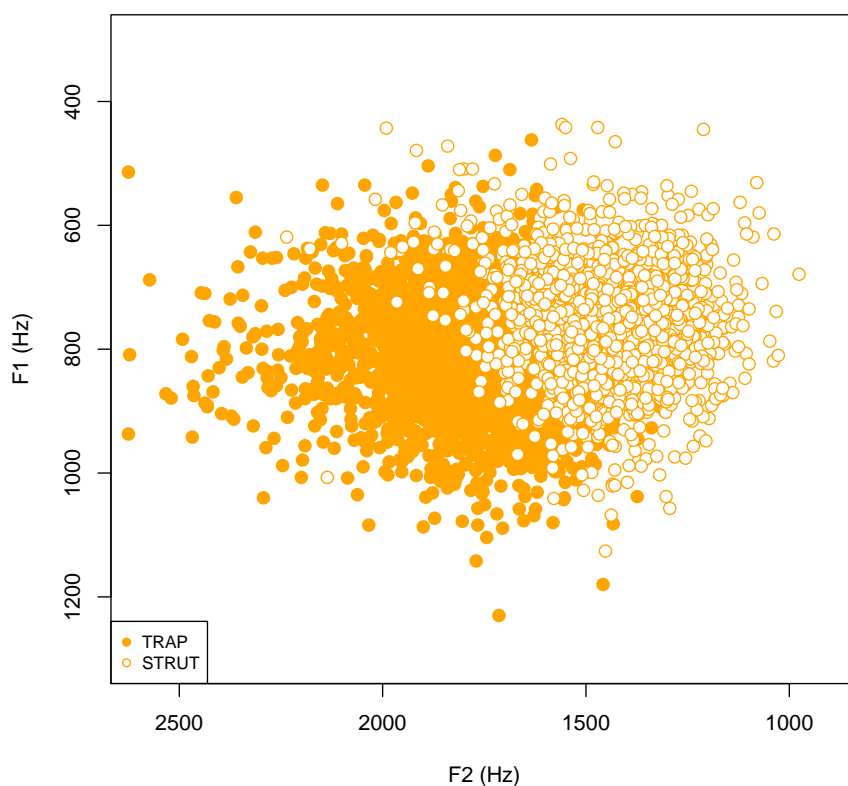


Figure 5.22: Casual style tokens of TRAP and STRUT with Pillai Scores between 0.5 and 0.8.

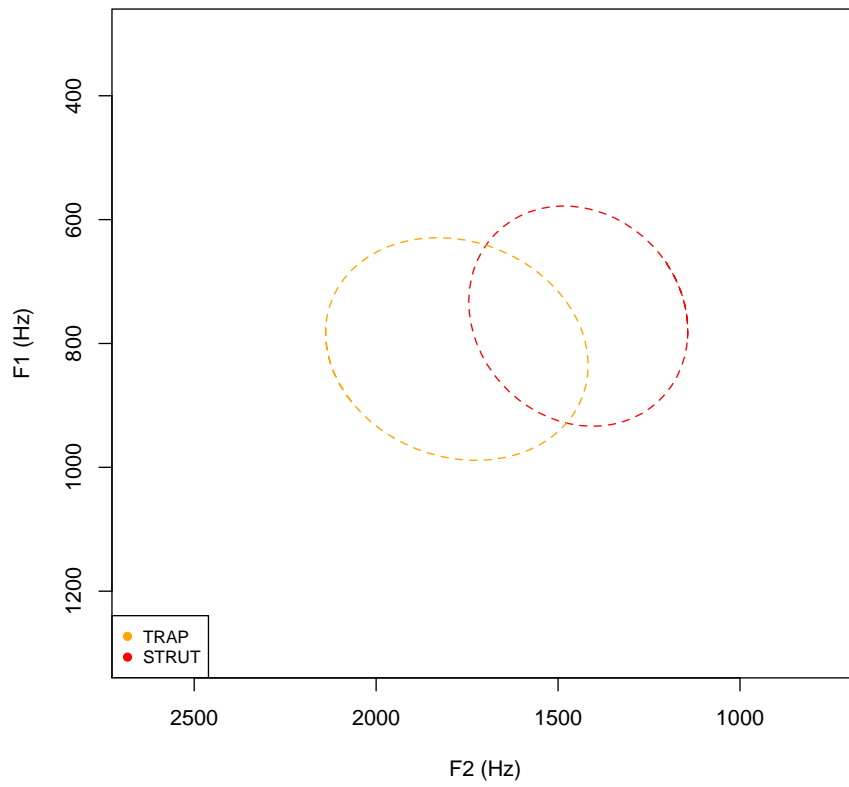


Figure 5.23: Standard deviation ($\times 2$) of all casual style tokens of TRAP and STRUT with Pillai Scores between 0.5 and 0.8.

5.6.1 MINIMAL PAIRS TESTS

Given the lack of merger in the casual style tokens, with only 21 speakers showing ‘close’ TRAP and STRUT, the minimal pairs investigation focuses only on these 21 speakers. The conditional inference trees for F1 and F2 (in figure 5.24) show that TRAP and STRUT realised in minimal pairs are significantly different from one another, even though there is some overlap. Visually, TRAP and STRUT are largely separate, as the formant plot in figure 5.25 shows.

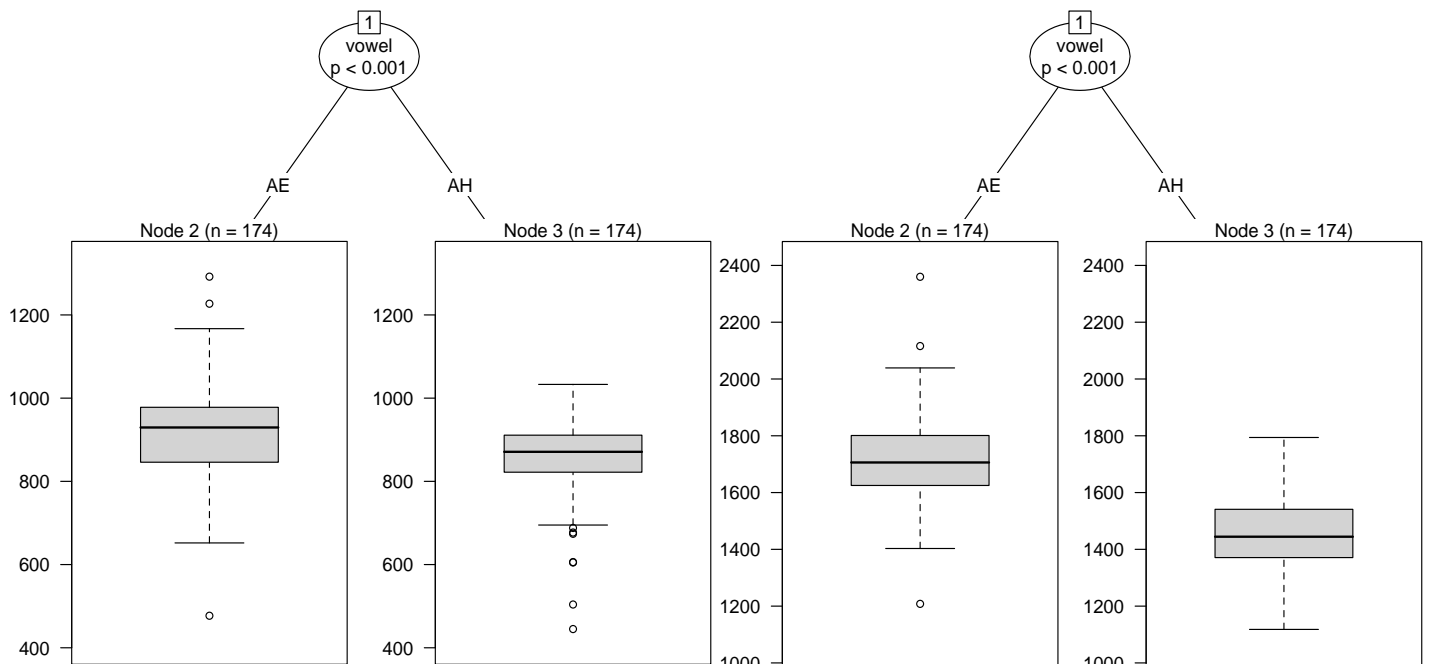


Figure 5.24: Conditional inference trees: TRAP and STRUT for F1 and F2 with ‘vowel’ as the predictor.

Each participant was asked whether the words in the minimal pairs test sounded alike or different. In each interview, a strong difference was articulated for each set. Not one of the 44 speakers heard these vowels as the same, and most of them produced a difference. Some speakers produced words similarly, most often ‘rant’ and ‘runt’, though these speakers were in the minority. Two such speakers were PerryF and LaurenL. Though these words sounded very similar, they return high Pillai Scores (0.73 and 0.60 respectively) and the distance is significant in each case ($p < 0.000$ for both). TRAP in most instances is lower and fronter than STRUT, also evident in figure 5.25.

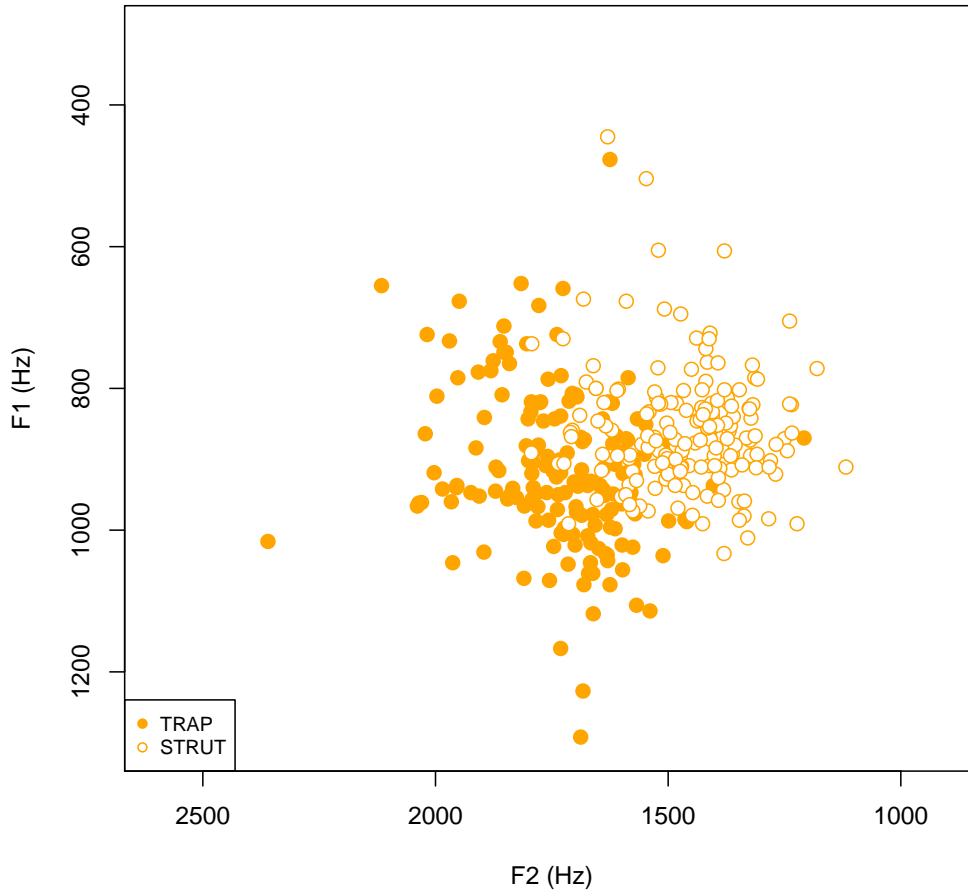


Figure 5.25: Minimal pairs of TRAP and STRUT tokens for 21 selected speakers.

Where all 21 speakers had Pillai Scores of under 0.5 for the casual style tokens of TRAP and STRUT, for the minimal pairs test the distance increased for most to over 0.5 (table 5.8). There is thus a greater distance between the vowels in the more formal style. Only three speakers remained well under 0.5: EricaV, JaneK and JennyL. Of these, only JaneK does not produce a significant difference between the vowels ($p=0.125$). This suggests that she does not distinguish between TRAP and STRUT in her production of minimal pairs, though she perceives them as different. Though the only speaker to behave this way, she shows a germ of truth in the claim that some SAfE speakers have overlapping TRAP and STRUT.

Speaker	Mean	TRAP	STRUT	Euclidean Distance	<i>t</i> -score	<i>p</i> Welch	Pillai Score	<i>p</i>
BobbyC	F1	896	895	368	0.052	0.959	0.888	0.000
	F2	1894	1526		11.883	0.000		
BradP	F1	930	851	339	4.992	0.000	0.643	0.000
	F2	1710	1380		4.438	0.001		
ChloëV	F1	938	844	194	2.724	0.018	0.675	0.002
	F2	1603	1433		3.151	0.014		
EricaV	F1	848	940	286	-1.205	0.2702	0.494	0.024
	F2	1758	1487		3.0804	0.010		
GaryL	F1	1125	943	399	3.872	0.004	0.912	0.000
	F2	1656	1300		11.17	0.000		
HayleyS	F1	894	868	337	1.082	0.297	0.660	0.000
	F2	1757	1421		5.798	0.000		
IlseP	F1	995	895	311	3.856	0.001	0.657	0.000
	F2	1699	1405		4.600	0.001		
JaneK	F1	982	930	247	1.375	0.207	0.371	0.125
	F2	1738	1497		2.405	0.049		
JennyL	F1	988	976	227	0.511	0.624	0.432	0.000
	F2	1602	1376		6.383	0.000		
KaleyW	F1	872	774	306	2.810	0.010	0.634	0.000
	F2	1675	1385		5.890	0.000		
KaraL	F1	896	843	273	0.747	0.472	0.596	0.010
	F2	1763	1495		3.901	0.004		
LarryR	F1	994	873	494	5.871	0.000	0.793	0.000
	F2	1925	1445		5.328	0.001		
LauraE	F1	914	901	243	0.365	0.719	0.597	0.000
	F2	1750	1507		5.038	0.000		
LeeL	F1	983	848	233	3.672	0.003	0.696	0.000
	F2	1673	1483		3.328	0.004		
LexiK	F1	934	845	299	2.745	0.014	0.654	0.000
	F2	1613	1328		4.956	0.000		
MindyG	F1	948	840	157	1.825	0.109	0.500	0.022
	F2	1704	1591		3.147	0.009		
PerryF	F1	722	814	241	-3.088	0.008	0.731	0.000
	F2	1938	1716		6.470	0.000		
RichS	F1	892	884	361	0.288	0.777	0.843	0.000
	F2	1742	1381		9.418	0.000		
RuthW	F1	900	837	186	2.111	0.049	0.674	0.000
	F2	1609	1434		6.031	0.000		
SarahN	F1	921	747	273	2.521	0.030	0.751	0.000
	F2	1741	1531		3.939	0.002		
TodK	F1	799	780	241	0.957	0.351	0.744	0.000
	F2	1754	1514		7.481	0.000		

Table 5.8: Pillai Scores and Euclidean Distances for TRAP/STRUT minimal pairs.

In conjunction with the results for the casual style tokens, it is clear that merger is not in progress despite the overlap. The individual vowel results presented earlier show that TRAP is backing, but also that STRUT is backing. They are repositioning, though not as Bekker (2009) foresaw. Instead of the upwards and frontwards movement of STRUT, it is an upwards and backwards movement. In this sense, STRUT can firmly be posited as part of the Reverse Vowel Shift in operation in SAfE, as it is in the Southeast English Chain Shift in Ashford (Torgersen & Kerswill 2004, 40).

5.7 CHANGES TO SOUTH AFRICAN ENGLISH

The discussion and analyses presented in this and the previous chapter have brought to light various on-going changes to South African English. Figure 5.26 is a broad schematic representation of these changes.

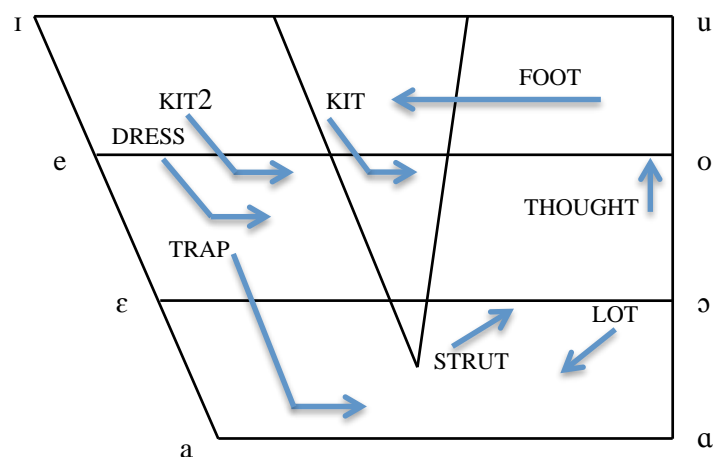


Figure 5.26: Vowel changes in operation in South African English.

There is concrete evidence of a reversal of the Older Vowel Shift in the lowering of KIT2, DRESS and TRAP (though it needs to be further borne out in the analysis of a bigger sample of older speakers than the ‘control group’ of this study). The KIT split is maintained by the secondary retraction of ‘other’ KIT, in response to the retraction of KIT2, an innovative change. TRAP lowers and retracts. This pushes STRUT upwards and backwards, and pulls DRESS downwards. KIT simultaneously lowers under pressure from the fronted FOOT vowel in the free position. The remaining vowels have a less certain part to play in the Reverse Vowel Shift, though THOUGHT raises and LOT fronts and lowers. The direction of the movement of LOT indicates that it is not part of the Reverse Vowel Shift. These changes are considered in more detail in the following chapter, which is, among other things, a comparison between SAfE and trends observed in other varieties of L1 English.

CHAPTER 6: DISCUSSION AND CONCLUDING REMARKS

This chapter summarises the findings presented in Chapters 4 and 5, whilst comparing SAfE to global varieties of English. It begins with a discussion of short front vowel lowering in section 6.1, paying particular attention to the possible causes of the Reverse Vowel Shift in relation to global vowel chain shifts and internal factors. Section 6.2 discusses the transmission of the new vowel shift by introducing a small pilot study. Section 6.3 considers what the changes to SAfE might index on a social level, while section 6.4 presents a summary of the findings in relation to the various research questions that this study set out to answer. In section 6.5 I discuss the impact of the research and its findings. Section 6.6 outlines how future studies can answer various questions raised by the analyses, after which I conclude.

6.1 SHORT FRONT VOWEL LOWERING DISCUSSED

The previous chapters have provided evidence for the undoing and reversal of the Older Vowel Shift in SAfE, manifest in the lowering of the short front vowels KIT, DRESS and TRAP. The FOOT and TRAP vowels are responsible for the Reverse Vowel Shift, which consists of both push and pull chains. The fronting of the FOOT vowel causes KIT to lower, while the extreme lowering of TRAP pulls DRESS downwards. Lowered TRAP also retracts, leading to overlap with the fairly front STRUT vowel. To avoid merger with the lax TRAP allophones, STRUT raises and retracts, and in this way forms part of the reverse chain shift currently in operation in SAfE. The retraction of KIT2 is an innovative feature, and though the secondary backing of ‘other’ KIT is associated with the continuation of the Older Vowel Shift, it is an important change in the Reverse Vowel Shift, leading to the maintenance of the KIT split. The retraction of DRESS has been noted as variable by previous scholars, and its stabilisation in younger speakers as a retracted vowel could be seen as a later generalisation of the retraction of KIT and TRAP, similar to KIT retraction in the Northern Cities Shift, a later generalisation of the backing of DRESS (Labov 2010, 112).

As outlined in Chapter 1 section 1.3.1, there are two criteria regarding causality that vowel changes must meet in order to prove the existence of a chain shift in progress.

1. Preservation of distinctions between the sounds involved.

2. Relatedness of the sounds in the following dimensions: temporal, spatial, and the speech of the individual.

The findings regarding the potential merger between TRAP and STRUT (Chapter 5, section 5.6) provide concrete evidence that the vowels shift in order to maintain distinctions. STRUT moves upwards and backwards, out of the way of TRAP as it lowers and retracts. Though there is overlap between KIT and DRESS, these sets both lower and retract, and the distinctions between the two classes are maintained. Moreover, the simultaneous lowering of KIT and DRESS are causally unrelated, since KIT lowers to avoid merger with FOOT, while DRESS lowers in response to TRAP lowering.

Temporal relatedness between changing sounds implies chronological ordering of the changes. To prove the existence of a vowel chain shift, therefore, is to show that one vowel changed first, leading to the subsequent change in other vowels. Regarding the Reverse Vowel Shift, the centralising of FOOT was the first change noticed by scholars on SAfE (such as Lass 1990, 277), and is therefore a plausible reason for the subsequent lowering of KIT. The lowering of TRAP was noticed slightly later (Bekker & Eley 2007; Bekker 2009), and can be cast as the impetus for later DRESS lowering. The findings of this thesis show that KIT and DRESS are in the process of lowering in response to changes in related vowels, a push from FOOT and a pull from TRAP, which is concrete evidence of chronological vowel chain shifting.

To solidify claims of a vowel chain shift it must be shown, in addition to temporal relations, that the sounds involved are related spatially, that is to say geographically and linguistically. Geographical relatedness requires the changes under investigation to occur within the same regional or social dialect. The first shard of acoustic evidence for the incipient reversal of the Older Vowel Shift was found in the work of Bekker and Eley (2007) and Bekker (2009), who noticed the lowering of TRAP in the citation style speech of young, white, middle class women. These findings are corroborated by my casual style findings, and are extended to the speech of young white men. Mesthrie (2012a) notes that the lowering of KIT, DRESS and TRAP is a big city phenomenon, which he has not noticed in his fieldwork in smaller centres such as Kimberley or Port Elizabeth. The Reverse Vowel Shift is therefore (currently) geographically located in South Africa's big cities, and socially is a feature of young, middle class, white speakers of SAfE.

The linguistic relation between changing sounds involves shared phonetic or phonological features. Ergo, sounds that are related linguistically are likely to move in a chain shift when the movement of one sound threatens the distinction between it and a related sound. There is proven relatedness and causality between TRAP and STRUT. The lowering of DRESS is a logical successive change, in that related vowels (in this case short front vowels) generally tend towards maximal dispersion in the vowel space (see for example Labov 2010, 95). In addition, the very fronted FOOT encourages the lowering of KIT to avoid the merger between these lax sounds.

The validity of considering the role of the individual in language change is subject to debate, though (Gordon 2002), argues that it is an important aspect to consider. If the Reverse Vowel Shift is a feature of young South African speakers of English, it follows that the causally linked sound changes should co-occur in the speech of the individual. For example, a young participant in the shift would not lower and retract TRAP without also lowering and retracting DRESS and raising and retracting STRUT. The majority of the young speakers, albeit to differing degrees based on gender, participate in all the changes attributed to the reversal and undoing of the Older Vowel Shift in the lowering and retracting of the short front vowels. Frequent outliers are MarilynK and CindyH, born in 1987 and 1995 respectively. I cannot postulate concrete reasons for the fact that they pattern with the speech of the older women at times, except that variation within a dataset is expected. Given the fact that there are only two outliers, the Reverse Vowel Shift can be posited as occurring in the speech of most young SAfE speakers, both individually and across individuals. In addition to the temporal and relational evidence, the existence of Reverse Vowel Shift is therefore also supported by the fact that it is evidenced in the speech of the individual.

Another important aspect to consider is whether the chain shift is phonetic or phonological change. According to Labov (2010, 290), chain shifting generally “maintains the unity of a segmental phoneme, while conditioned sound changes exhibit the disruptive effects of coarticulation.” Though there is variation based on the phonological environments in which each vowel occurs, the overall tendency is lowering and retraction, and there is no evidence

to suggest that any allophone has ceased to be a member of the phoneme.⁷³ This suggests that the phonemes KIT, DRESS and TRAP undergo phonetic change.

Variation based on environment is perhaps evidence of change in progress yet to stabilise. An example of this can be found in the DRESS data. Generally, the younger the speaker, the lower the vowel. Men and women between 24 and 32 years of age lower DRESS before /m/ and /n/ more than the other environments. Men and women between 18 and 23 years of age lower DRESS equally before /m, n, k, ŋ/. Though this is a very fine-grained age result, the increase in environments displaying lowering shows that the phonetic lowering of the phoneme is progressive. It is expected, therefore, that the remaining environments (the free environment, and DRESS before /g/), though also lowered in apparent time, will have lowered to the same position as DRESS before /m, n, k, ŋ/ when the change has stabilised.

Conditioned or allophonic shifting involves the shifting of vowels in very specific environments, and not at all in others (Labov 2010, 292ff). In other words, if lowering and retraction were constrained to the velar environments, for example, this would be allophonic shifting, and the unity of the phoneme would be disrupted. The lowering and retraction as described for my dataset shows no such phonological constraints. The argument for phonetic change to the entire phoneme is furthermore supported by the fact that the younger speakers (the shifters) are often statistically different from the non-shifters or the older speakers. This means that, in general, they realise these vowels statistically differently as a phonological whole, not on an allophonic basis. The segmental phoneme remains intact, and is not disordered by conditioned change.

Though it is still early on in this shift, particularly for KIT and DRESS, the findings suggest that the entire phoneme (encompassing all the allophones) for each vowel will shift downwards and backwards phonetically. In the abstract, underlying phonological system, however, the phonemes do not shift. The underlying elements of each phoneme have not changed: the lexical sets are ‘intact’ in that they have not merged with other lexical sets. Though there is a change in phonetic realisation, phonetic distance is maintained. In this sense, though the phonemes and their allophones undergo the Reverse Vowel Shift, the change is not phonological.

⁷³ Though DRESS_1 is much lower and backer than the other DRESS allophones, it is still unequivocally part of the DRESS set.

6.1.1 ENDOGENY OR EXOGENY?

An important socio-historical question is whether the undoing of the older shift is caused by endogenous (internal) or exogenous (external) factors. The Older Vowel Shift, following Labov's Pattern 4 as part of the Southern Shift, is posited as an endogenous change stemming from the British settler input varieties. Lass and Wright (1986) see it as a chain shift when compared to the input varieties, whereas Trudgill (2004) argues that the raised vowels are merely a conservative form of southern British Englishes, which began to lower the short front vowels. In the same vein, Trudgill, Gordon and Lewis (1998, 49) suggest that the raised short front vowels in New Zealand English are 'inherited from' the settler input British English dialects, which they support by comparisons to (older) SAfE and Australian English:

This conclusion is of course greatly strengthened by the fact that Australian and South African English also have much closer vowels than those of English English: the occurrence of a single innovation — that of lowering in England — is very much more likely to be the correct explanation for this differentiation than the occurrence of three separate but identical innovations that just happened to take place at about the same time in three different and widely separated parts of the world.

Whether the raised short vowels characteristic of older SAfE (and other Southern Hemisphere Englishes) results from a chain shift or are conservative dialects, the underlying motive for their comparative closeness is certainly, therefore, endogenous. Australian English and SAfE have recently begun lowering their short front vowels. New Zealand English is the only Southern Hemisphere English variety that does not experience this lowering, but which instead underwent secondary raising of KIT, DRESS and TRAP (Trudgill 2004). If raised vowel qualities are endogenous features, it is important to consider the motivations for their recent lowering, particularly in SAfE.

The argument for the endogeny of the Reverse Vowel Shift would contend the following:

1. Short front vowel lowering is a common occurrence in languages around the world, explained via Labov's Principle II: lax nuclei fall along a non-peripheral track.
2. Short front vowel lowering occurred in southern British Englishes long before it occurred in the Southern Hemisphere Englishes (beginning during the mid 1900s), which contributes to the differences between British Englishes and the Southern Hemisphere varieties. The recent lowering of the short front vowels in Australia and

South Africa could thus be explained as colonial lag (as discussed in section 2.1 of Chapter 2).

Considering how current SAfE can be explained via Labov's Principles and Patterns, the results presented in Chapters 4 and 5 suggest that SAfE in Cape Town currently experiences Pattern 2 chain shifting. This involves the fronting and raising of long vowels (Principles I and III, also subsumed in Pattern 3) with falling short vowels (Principle II).

The GOOSE vowel is well described as showing extreme fronting in SAfE (see for instance Mesthrie 2010). This fronting has been described as an endogenous, Pattern 3-like change, exemplifying Trudgill's notion of 'drift' (Bekker 2009, 313; Chapter 2, section 2.1). FOOT has been described as centralising in SAfE as early as Lass (1990), corroborated by the data presented in Chapter 5 in addition to evidence of further fronting by younger speakers. Bekker (2009, 359ff) notes that the fronting of FOOT is an endogenous change from below, operating generally under Principle III. Though my findings are preliminary, THOUGHT seemingly undergoes (secondary) endogenous Pattern 3-like raising (Bekker 2009, 312) along the same lines as which GOOSE fronts. FLEECE appears to be fronting even further. In conjunction with short front vowel lowering, it is therefore likely that SAfE is currently undergoing Pattern 2 chain shifting.

Through the general adherence to the internal motivations for chain shifting that Labov (1994) sets out, the Reverse Vowel Shift together with the various other changes are likely internally motivated, particularly given the fact that numerous other Englishes show the same patterns. This argument is further strengthened by Labov's (2010, 186) observation that the Canadian Shift, *inter alia*, occurs below the level of social consciousness, and is thus an internally motivated change not influenced by external factors such as language contact.

Another explanation for the endogenous nature of the changes that culminate in the Reverse Vowel Shift is very possibly colonial lag. Bekker (2009, 204-6) argues that TRAP lowering is likely an endogenous change as a result of a delay in linguistic change, in that SAfE and Australian English mirror the lowering that has already occurred in RP and in southeast England. The lowering of TRAP in RP began in the mid 1900s, reported on as early as 1962 by Gimson, and later by Roach (1983) and Fabricius (2007). Short front vowel lowering was noticed in the southeast of England as early as Wells (1982, 129), instigated by the earlier lowering of TRAP (Trudgill 2004). It therefore seems plausible that the later lowering of

particularly TRAP, starting in the speech of South Africans born around the year 1982, is a result of colonial lag.

Bekker (2009, 206) states that if the lowering of TRAP can be conclusively shown as an endogenous, structurally motivated change, the evidence would support an incipient chain shift like the Southeast English Chain Shift. In support of this claim, he adds that the lowering of TRAP appears to be prestige driven, moving away from the raised TRAP stereotypically associated with Broad SAfE. In this sense, he suggests that short front vowel lowering in SAfE is a change from above. Maclagan and Hay (2007, 40) posit similar findings for TRAP lowering in Australian English, concluding that it is prestige driven and potential hypercorrection.

Bekker's (2009, 433) overall findings for SAfE support the notion of an incipient vowel chain shift akin to what Torgersen and Kerswill (2004) describe. He ascribes these changes largely to colonial lag, especially in that his findings show that SAfE remains 'conservative' in the fronting of LOT and STRUT. Though LOT remains a 'fronting' vowel in my findings, STRUT is retracting in response to TRAP backing. This is another similarity to the vowel changes present in the Southeast English Chain Shift, supporting an endogenous and colonial lag explanation for the Reverse Vowel Shift.

Mesthrie (2012a, 2098) supports the notion that the lowering of the short front vowels could be a change from above, particularly since he noticed the incipient change as having a "middle-class, prestigious, experimental, or innovative aspect to it." Mesthrie (2012a, 2099) however also considers exogenous origins of the change. The first possible solution he offers results from what he calls "recent social history." The time when the lowering became more frequently noticed (around the year 2000) is a time associated with socio-political change as well as the 'gap year' phenomenon, as mentioned in Chapter 2. The lowering of particularly TRAP and KIT could be as a result of diffusion, given the fact that many young South Africans were in direct contact with dialects in the UK during their gap year(s), and then potentially brought the feature with them upon their return. The second possible solution Mesthrie (2012a, 2099) offers, given the parallel short front vowel lowering observed in Canada and Australia, is that these changes are all in response to an earlier prime mover, which he contends is California via the popularity of Hollywood.

The English-language programming on South African television does not feature much locally produced content when compared to content imported from the United States and Britain. South African television, particularly satellite television (DSTV), showcases more American than British productions, particularly for children (especially formative educational programmes like *Barney* and perhaps *Sesame Street*), but also for adults. Much of the audio-visual media consumed by South Africans therefore leads to regular ‘virtual contact’ with American Englishes. In observing my preadolescent nieces at play, I have noticed that they always use ‘American’ accents when they role-play, never ‘British’ accents, indicating an association between ‘acting’ and ‘American’. In a recent primary school production of *Annie*, everyone was clearly directed to speak with an ‘American accent’. The children did this consistently and confidently, though the adult teachers in the play either did not, or struggled to maintain it. This either suggests great familiarity with American English on the part of the children, that confidence is key, or both.

Not only is much of the television programming in English imported from the United States, many adverts on the radio appropriate American accents. KFM and 5fm are popular South African radio stations. 5fm broadcasts countrywide, and is particularly associated with young people. From personal experience, the closer one gets to 30, the more appealing KFM, the local Cape Town station, becomes as an option for someone slightly older than the market 5fm appeals to. On both stations, one regularly hears American-like accents. KFM in particular broadcasts a late-night show hosted by Ryan Seacrest, a popular American personality in radio and television. Music is also largely American, in that many artists are from America, in addition to many (though by no means all) non-American artists who sing with American-like accents. For example, the Australian boy band *5 Seconds of Summer* sound American in their songs. Mick Jagger from *The Rolling Stones* has long sung with a largely American accent despite his British nationality. *Prime Circle* is a popular South African band whose lead singer sounds more American when he sings, despite his British heritage. In this way, much of the media consumed by South Africans leads to regular hearing of American or American-like Englishes.

Hickey (2015, personal communication) posits that the short front vowel lowering in Dublin is influenced by America, which is seen as a source of ‘coolness’. England is not seen this way due to the historical relationship between it and Ireland. Hickey furthermore notes that Americanisms creep into Dublin English through the media, citing short front vowel lowering

and retroflex /r/ as examples. Though the relationship between Britain and South Africa is not quite the same as between Britain and Ireland, it would appear that there is indeed a remarkable American influence on English in South Africa. For one, it would appear that America is seen as appealing on a global scale: in a discussion with the second year students in Linguistics at UCT, they noted the appeal of sounding ‘American’ in that it resonates with global awareness in a way which the older SAfE accents do not. Moreover, there appears to be an increase in rhoticity (Hartmann & Zerbian 2009; Du Plessis & Bekker 2014). Though not dealt with in this study, rhoticity is certainly a feature for some of my speakers, mostly in the word list for items like *bird* and *nurse*. This feature of neorhoticity was also found in Du Plessis and Bekker’s (2014) study. An interesting case from my database is TedR, who was born in 1994. He is largely a rhotic speaker in casual style, without having been in America or any other place where the English is rhotic. Interestingly, his parents ‘blame’ his *r*-fullness on the many hours he spent watching television as a child.

The Reverse Vowel Shift mirrors the shifts that have been described for California and Canada in particular, though it also shares important features with the Southeast English Chain Shift in Britain (DRESS and TRAP lowering; STRUT backing; FOOT fronting), which could indicate colonial lag. Given the dominance of Hollywood in English-language broadcasting in South Africa, it seems very likely that virtual contact with American, particularly Californian, English could be responsible for TRAP lowering, which in conjunction with endogenous FOOT fronting, instigates the vowel chain shift this thesis attests for SAfE in Cape Town. Other varieties of English also undergo short front vowel lowering, particularly those in Dublin and Australia. The Dublin English chain shift is influenced by American English (Hickey 2013), and it seems reasonable to posit colonial lag and/or virtual contact as the reason for the Australian vowel lowering. Given vowel raising once characteristic of SAfE, there was ample room in the vowel space to accommodate lowering vowels without merger. Considering these arguments, it seems likely that the Reverse Vowel Shift in SAfE is an externally motivated change.

If however, the notion of virtual contact is something considered impossible, as it is by Chambers (1998), Mesthrie (2012a, 2100) suggests that a ‘multiple influence scenario’ may have to be accepted. In this sense, the Reverse Vowel Shift in SAfE may very well illustrate colonial lag, and can be seen as a change from above. In explaining the possible results of the glide weakening of fronted PRICE in General SAfE, Bekker (2009, 188) uses Mufwene’s

(2008) theoretical approach to the ‘ecology of a variety’ as a solution, which considers both internal and external factors as important to the genesis of a language variety. Hence, the Reverse Vowel Shift is probably best described as an endogenous change showing colonial lag, but which has been given prominence or has been sped up by the virtual contact with certain features via the media as well as via direct contact by gap year takers. In other words, the endogenous fronting of FOOT is causing KIT to lower, which would likely have caused DRESS and TRAP to eventually lower in an endogenous push chain. TRAP, however, has already lowered and retracted in SAfE. This rapid progression (in a way skipping the DRESS lowering stage) suggests that it was motivated by more than internal factors. SAfE has the propensity to lower TRAP in the sense of colonial lag, and virtual contact with TRAP-lowering dialects of particularly American, but also British, Englishes, speeds up the process, causing DRESS and KIT to lower simultaneously.

6.2 TRANSMISSION OF THE REVERSE VOWEL SHIFT

Linked to the stabilisation and spread of the change, issues surrounding diffusion and transmission are important to consider. When adults participate in language change due to contact, they learn the general feature and apply it in a way that is different from the input variety. Labov (2010, 316ff) explains this via the short *a* (/æ/) system in English in New York City: speakers of the input dialect have specific constraints on the lengthening and raising of /æ/. As the feature spread to other areas within fairly close proximity, there is evidence of a loss of constraints. In Northern New Jersey, for example, speakers are found to lengthen the vowel in open syllables and in function words, both of which are constraints on lengthening in New York City (Labov 2010, 321). Adult language learners ‘hear’ the tensing and apply it without the constraints evident in the speech of a native user of the feature. In other words, adults adopt the change ‘imperfectly’. The children who come into contact with a particular feature will, in all probability, learn it with all the constraints intact, in which case the change is transmitted to children, and diffused to adults (see Labov 2010). Given the overall conclusion that the Reverse Vowel Shift is a feature of young adult speakers of SAfE in Cape Town, it is important to ascertain whether children participate, and if so, in what way.

A small pilot study to this end was carried out in the first academic semester of 2015 by a postgraduate sociophonetics class at UCT. Each of the students was tasked with interviewing a pre-adolescent girl between the age of 8 and 12. In addition, they were to be from Cape Town, middle class and white, to ensure comparability with my data. Using the same method of analysis as has been employed in this thesis the five young girls (born between 2002 and 2005) were compared to six women from my dataset who are all Reverse Vowel Shifters (born between 1992 and 1994).⁷⁴ Figure 6.1 is the scatter plot of the results, after I checked and recoded the data where necessary.⁷⁵

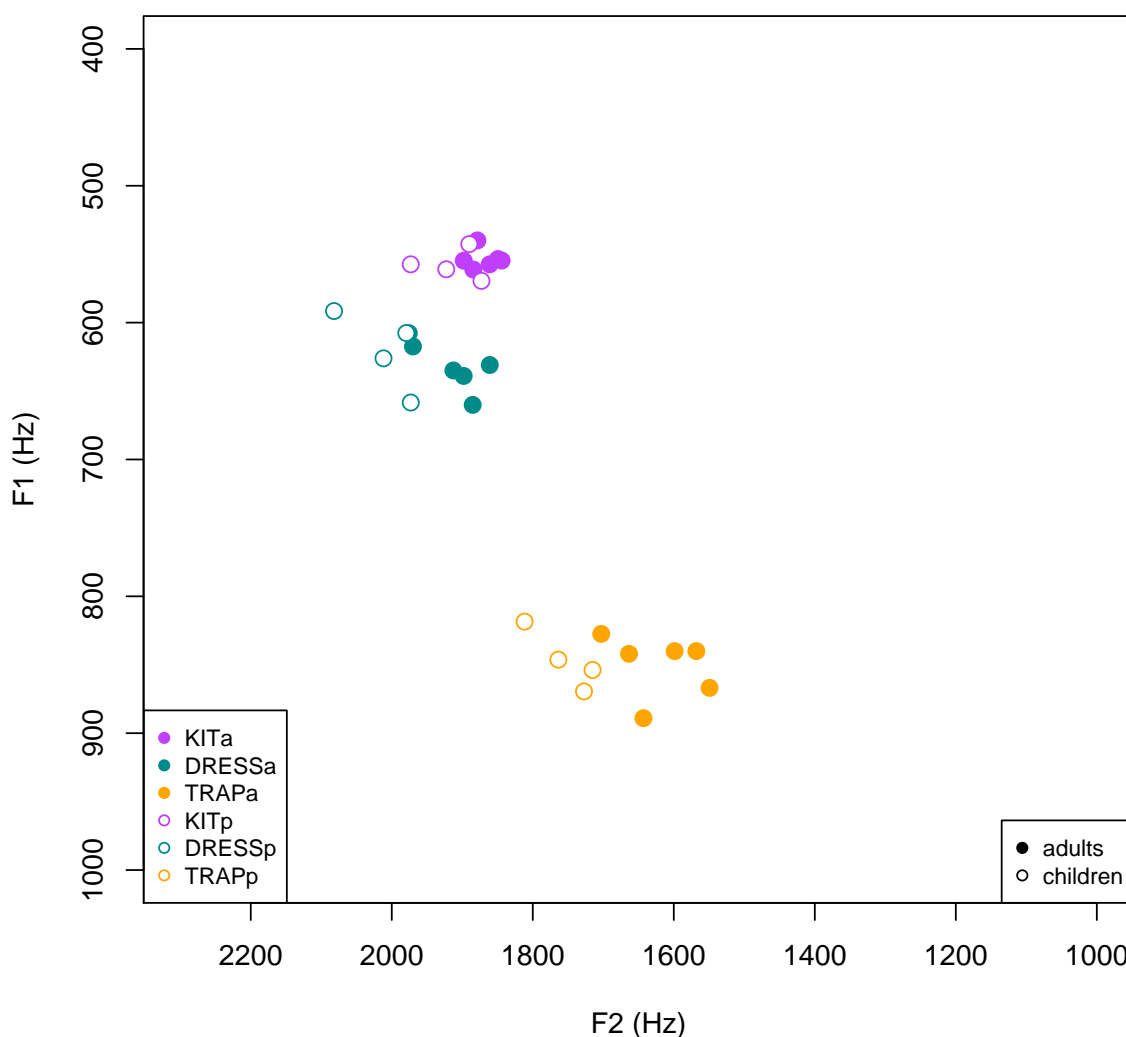


Figure 6.1: Comparison of 6 adult women (a) to 4 pre-adolescent girls (p) for KIT, DRESS and TRAP.

⁷⁴ Originally, it was intended that there be six children and six adults, but unavoidable circumstances resulted in only five children being interviewed and analysed.

⁷⁵ There was a continuous outlier who I have excluded here – hence there are only four children in the analyses to follow.

For KIT, a mixed-effects model suggests that ‘date of birth’ is not a significant predictor of F1 ($p=0.701$), but that it is for F2 ($p=0.016$). The results for DRESS and TRAP are the same: there is no age difference for height ($p=0.661$; $p=0.898$), though there is for frontness ($p=0.020$; $p=0.007$). In general, therefore, the children lower each vowel as much as the adult women, though retraction does not occur to the same extent.

A conditional inference tree for KIT (not produced here) supports the claim that the children lower the vowel as much as the adults do: there were no significant differentiations based on the predictors ‘age’ or ‘environment’. Regarding retraction (figure 6.2), age is not a significant predictor of F2 for KIT0, KIT5 and KIT6: both age groups treat these environments in the same way (node 5). When KIT occurs in the free environment (KIT0), it is generally fronter than KIT occurring around bilabials, after /r/ and /l/ and before /w/ (KIT5), or after /w/ and before /l/ (KIT6). This is expected given the retracting qualities of the environments encapsulated by KIT5 and KIT6. These three environments, however, are significantly retracted when compared to KIT2 and KIT3, which is where there is an age difference. The women retract the vowel in these two environments (node 3), where the children maintain generally fronter realisations (node 4). It would therefore appear that the children do not participate in the backing of KIT2 as the adults have been shown to do.

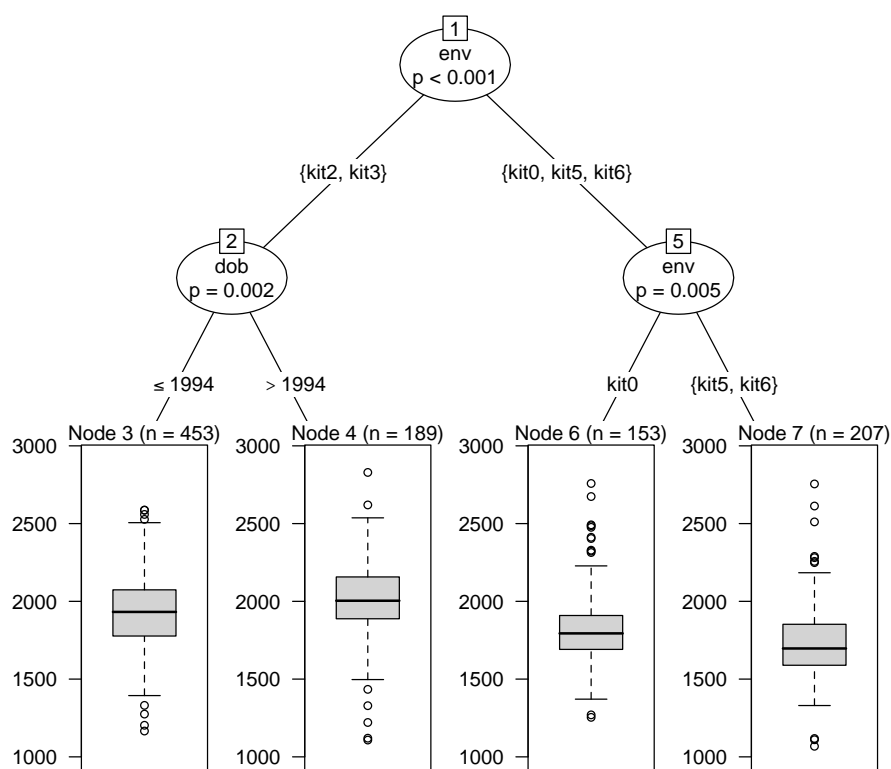


Figure 6.2: Conditional inference tree for KIT F2 by ‘environment’ and ‘date of birth’: children versus adults.

Interestingly, the grouping of the environments suggests that the speakers do not have a KIT split as shown in Chapter 4. However, if we analyse each age group individually, it appears that they both display a binary split between KIT2 and ‘other’ KIT. Figure 6.3 shows the split for the women, and figure 6.4 the split for the children.

Though both groups clearly have a KIT split in their speech, the women have a general centralising tendency, where the children have a binary split between frontier KIT2 and the remaining environments (‘other’ KIT). These ctree analyses further support the analysis presented in figure 6.2: the children do not retract KIT2 as much as the adults do. Where the adult median sits well under 2000 Hz (node 2, figure 6.3), the children’s median is just about 2000 Hz (node 2, figure 6.4). In conjunction with figure 6.1, it furthermore shows that the children retract ‘other’ KIT in the same way as the adults, despite the fact that the adults show some environmental variation. The disparity between figure 6.2 and figure 6.3 and 6.4 could be a result of a few factors: (a) there are only 21 KIT3 tokens in total, which might lead to the skewing of the data; and (b) there are 6 adults compared to 4 children, which could result in these odd results due to the skewness (or size) of the sample. A bigger and more balanced sample is required.

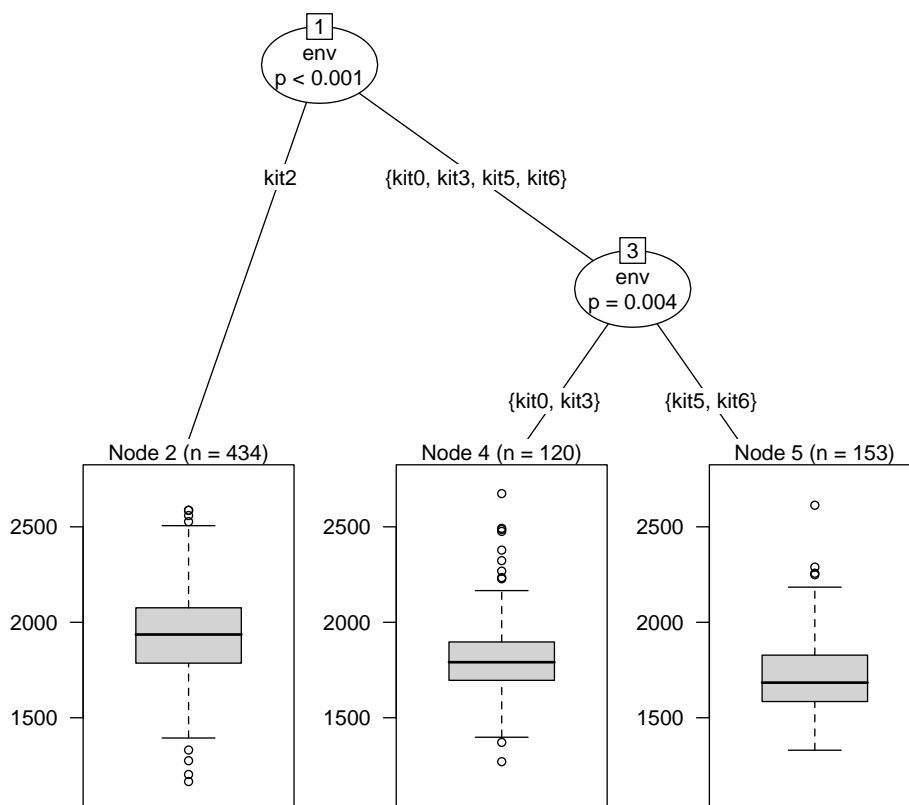


Figure 6.3: Conditional inference tree for six women, showing the KIT split along F2.

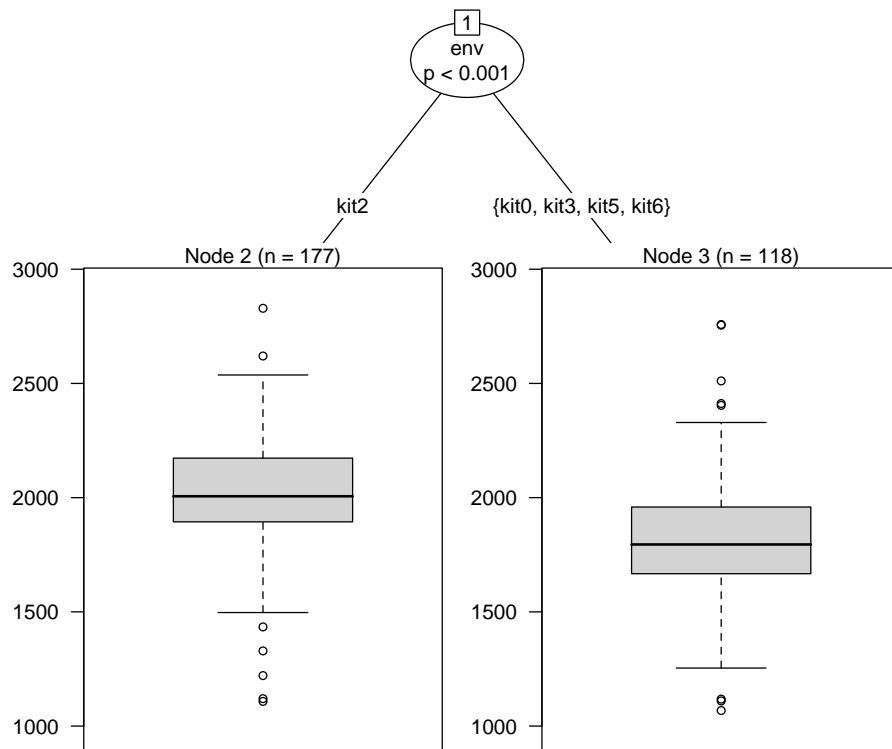


Figure 6.4: Conditional inference tree for four children, showing the KIT split along F2.

As the scatter plot in figure 6.1 suggests, there are varied levels of DRESS lowering by the children in the sample. Though there are children who lower in a way very similar to adults visually, a ctree analysis shows slight differences (figure 6.5). There is no difference according to age for DRESS_1 (node 2), DRESS0 or DRESS_g (node 7). It therefore appears as though the children participate in the lowering of the vowel, particularly in the environments DRESS_0 AND DRESS_g. The only age difference is for DRESS when it occurs before /k, m, n/, as is seen in node 4. The children (node 6) realise DRESS as a slightly higher vowel than the adult women, who lower the vowel to about 650 Hz (node 5). Generally speaking, the two groups produce DRESS around 600 Hz along F1. The fact that the adults lower DRESS in only three environments could account for the discrepancy between men and women in the lowering of DRESS, evidenced in Chapter 4, though this would require further investigation.

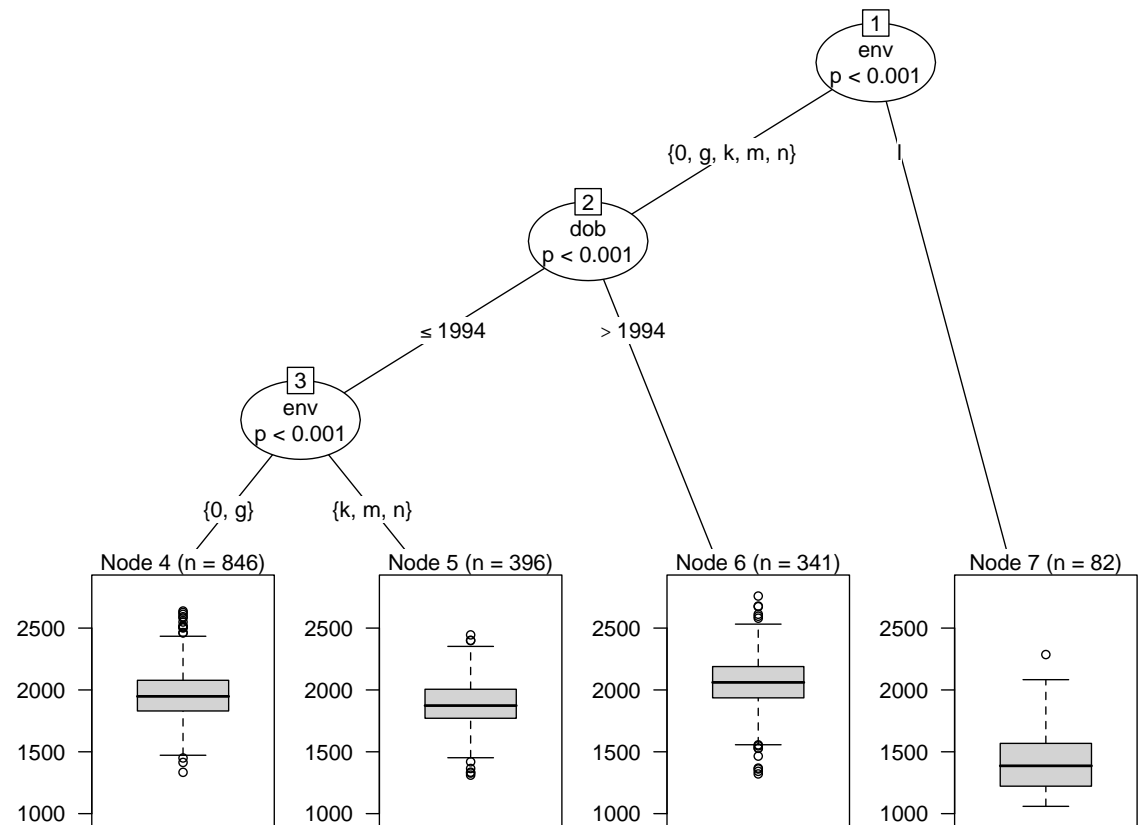
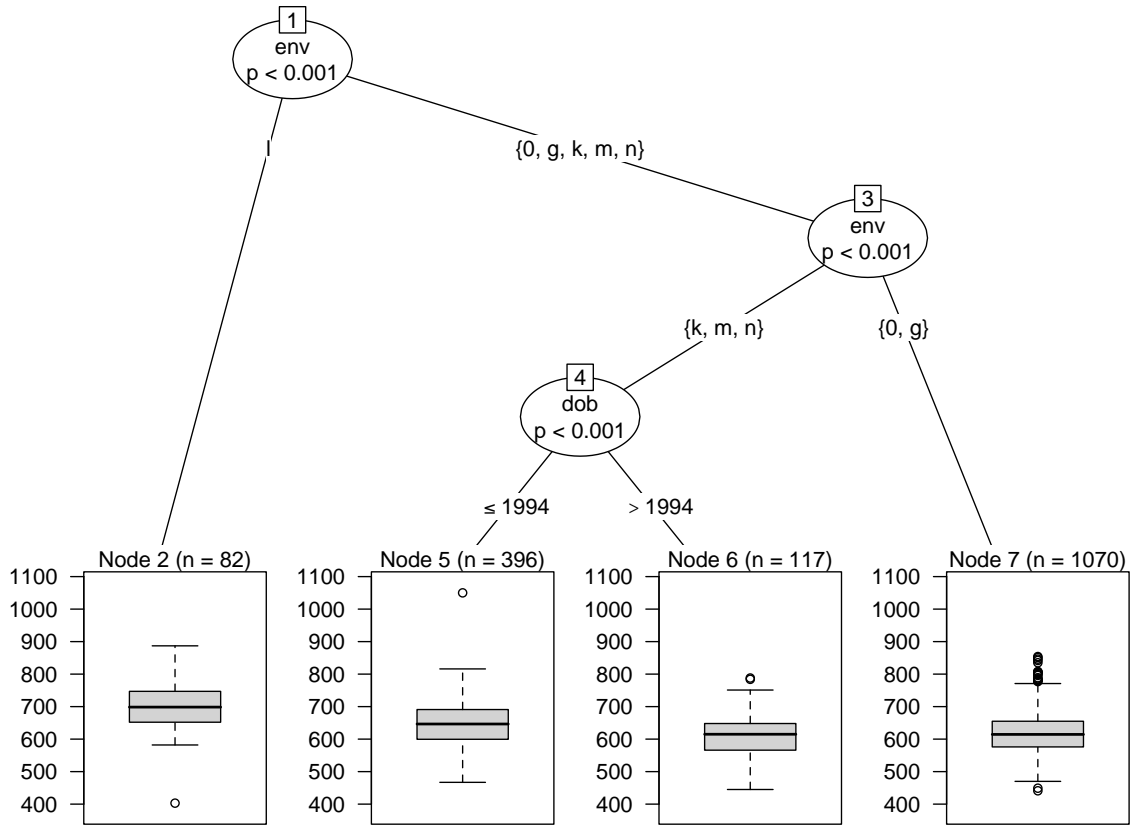


Figure 6.6 shows the comparison between the two groups for the F2 of DRESS. It shows that the children participate in the secondary backing of DRESS_1 (node 7). Where the adult women show some environmental distinctions, the young children do not, and produce DRESS uniformly at just above a 2000 Hz median along F2 (node 6). This is considerably more front than the women, who never reach a median 2000 Hz in their realisations (nodes 4 and 5). It therefore seems that the children do not retract DRESS, except when it occurs before /l/.

Regarding the lowering of the TRAP vowel, the children and the women lower the vowel in exactly the same way. The ctree analysis in figure 6.7 shows that age is not a significant predicting factor, and that TRAP is lowered the most in the free position, and before nasals.

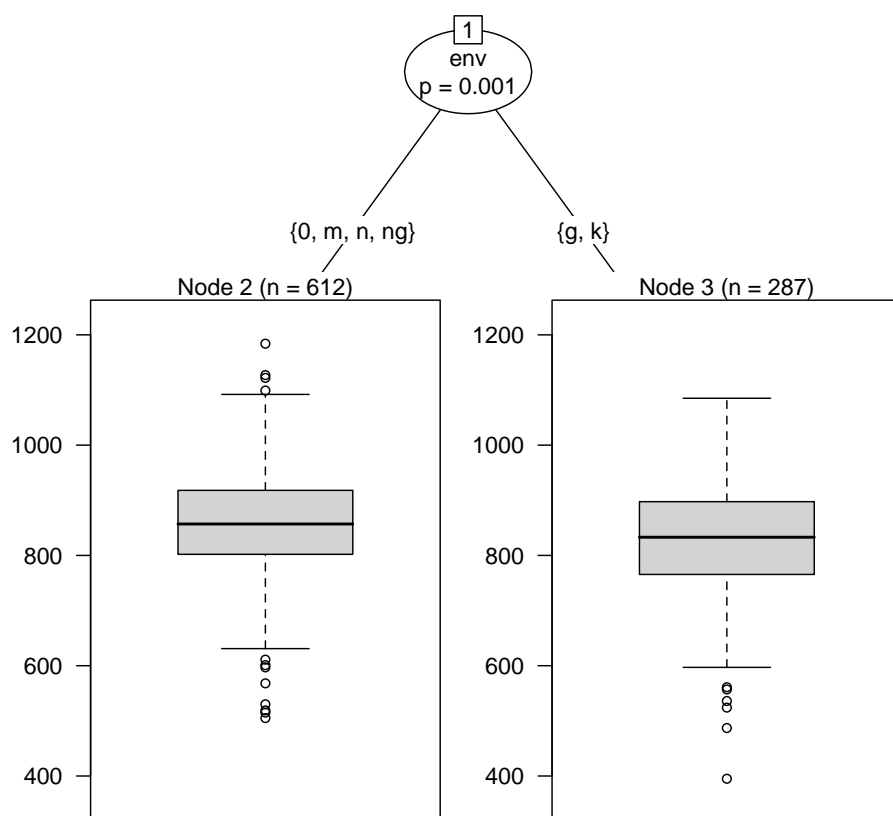


Figure 6.7: Conditional inference tree for TRAP F1 by ‘environment’ and ‘date of birth’: children versus adults.

The retraction of TRAP is not as neatly conditioned by age, as is evident in figure 6.8. There is some differentiation between the adult women: those born in 1992 or 1993 (nodes 3 and 4) differ from those born in 1994 (node 6). Conversely, the children (node 9) are (a) uniform in their realisations, both regarding exact age and environment; and (b) realise the TRAP vowel consistently more front than the older speakers, especially those born in 1992 or 1993. When we compare the children to the adults born in 1994, it would appear that they produce TRAP in the same way, particularly when it occurs before /k/ or /n/ in the women’s speech (compare notes 7 and 9). In comparison to the other environments (node 8), and with the other older women (node 2), the children’s realisations are generally fronter. Though there is variation in the level to which the young women retract, young women in general have been shown to retract TRAP (as per Chapter 4). Given the partial overlap between the children and some of the adults, it follows that the children also retract TRAP, but not as much as the older women.

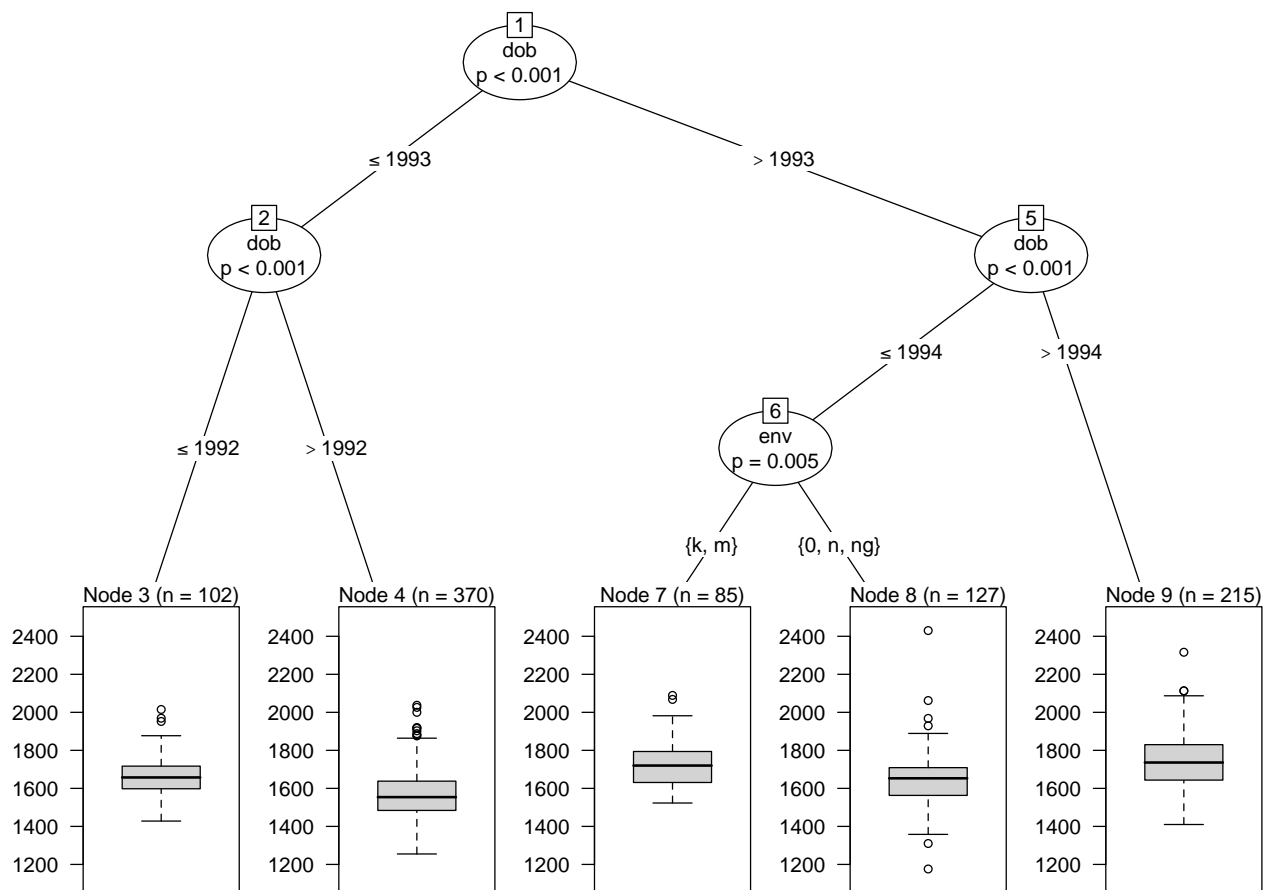


Figure 6.8: Conditional inference tree for TRAP F2 by ‘environment’ and ‘date of birth’: children versus adults.

In lowering KIT and TRAP, the children show participation in the Reverse Vowel Shift. DRESS is a complicated vowel, especially in young female speech, though in general there is overlap between the height of the vowel class between the age groups, especially in the free position and before /g/. In these environments, the children seemingly lower the vowel as much as the young women. The retraction of KIT exhibited by the adults is mirrored in the speech of the children except for KIT2, which is not retracted. This signals participation in the Reverse Vowel Shift in the sense that short front vowel retraction is a feature of the shift in addition to lowering. Though the children participate in the secondary retraction of DRESS_1, the remaining DRESS allophones are not retracted in the same way as by the adults. Similarly, TRAP is not as retracted in the children's speech than in the speech of the adults, though some retraction is evident, suggesting participation in this instance as well.

It would therefore appear that the children participate in the Reverse Vowel Shift by lowering KIT, DRESS and TRAP (with the situation surrounding DRESS requiring further investigation). The retraction evident in the speech of the women cannot be claimed as a feature of the children's speech conclusively, though some retraction is evident. This casts a somewhat complicated light on the matter of diffusion and transmission. The children are very clearly in contact with short front vowel lowering, be it via the media they consume or contact with slightly older peers, and they apply it variably in their English. At this stage, no clear inferences can be drawn about how the change is spreading through the community. Further research would illuminate the way in which children acquire the shift, which would in turn solidify issues pertaining to the transmission and stabilisation of the Reverse Vowel Shift in SAfE.

6.3 INDEXICALITY AND ENREGISTERMENT

Considering the indexical meanings behind the changes to the short front vowel system of SAfE could enrich our understanding of its origins and development. Eckert (2008) provides a very strong argument for expanding the focus of inquiry into language variation and change. In particular, Eckert (2008, 472) holds,

A theory of variation ultimately must deal with meaning, and not only does a view of meaning in variation as predetermined and static seriously undershoot human capacity, it cannot even account in any principled way for the changes in correlations that have been observed in the lifetime of a sound change.

That the Reverse Vowel Shift was on one hand motivated by influence from other Englishes also prone to short front vowel lowering is a likely reason for the adoption of the trend in SAfE. Equally, it must have subconscious social symbolism to become accepted and transmitted. Referring to the Northern Cities Shift, Eckert (2008, 454) notes that it would be “futile and ridiculous” to try to explain it in terms of the making of meaning in everyday interactions. However, she further states that the dismissal of what people do with elements of a chain shift to create meaning is to ignore an “aspect of human competence that is at least as mind-blowing as the ability to maintain distance between one’s vowels.” Similarly, Johnstone and Kiesling (2008) argue that a careful investigation of and approach to the creation of meaning via linguistic features yields a more nuanced understanding of language variation and social meaningfulness than can an account of production or perception alone.

These scholars all point to the usefulness of Silverstein’s (2003) work on indexical order. Silverstein (2003, 193) claims that, “‘indexical order’ is the concept necessary to showing us how to relate the micro-social and the macro-social frames of analysis of any sociolinguistic phenomenon.” In describing the different ways in which meaning is indexed by linguistic variables, three indexical ‘orders’ are posited. ‘First-order’ indexicality is similar to the term ‘indicator’ as described by Labov (1972), in that it is a pattern often noticed by linguists, and social meaning is ascribed to the variable based on what the patterning of its use tells us. Users of the feature are usually unaware of and cannot use it to consciously project or characterise social meaning (Johnstone & Kiesling 2008, 10). A first-order index situates a speaker within a general population, such as being a Martha’s Vineyarder by centralising the PRICE set (Eckert 2008, 463). Labov (2010, 186) supports this first-order correlation between

accent and place, specifically stating that the social indexes found in Martha's Vineyard are below awareness.

When a first-order index is used to create meaning in the sense of social distancing or association, it becomes a 'second-order' index, similar to Labov's 'marker'. People are aware of and use the particular linguistic variable, though they are not as a result able to discuss it meaningfully. Once a speaker is aware of and able to engage in a discussion or description of the variable that indexes a particular social characteristic, it has become a 'third-order' index, similar to Labov's 'stereotype' (Johnstone & Kiesling 2008, 10).⁷⁶ Continuing with the Martha's Vineyard example, the centralisation of PRICE would become a second-order index if members of the speech community use it consciously to index their loyalty (or the absence thereof) to the island. If the community were fully aware of and able to articulate the feature and its social meaning (for example in stand-up comedy or in other ways to explicitly characterise local 'Martha's Vineyardness') it would become a third-order index. A concrete example is MOUTH monophthongisation in Pittsburgh: the community is so aware of it that souvenirs such as t-shirts and coffee mugs represent it as a feature of 'Pittsburghese' (Johnstone & Kiesling 2008, 5).

Though my study did not set out to expressly investigate what the Reverse Vowel Shift indexes on a social level, general comments can be made about possible, particularly first-order indexes drawn from the analyses. Given the lack of qualitative work on the attitudes to or ideologies surrounding English in Cape Town, no substantiated claims can at this stage be made about possible second- or third-order indexes. It could of course be that reverse shifting does not index anything beyond the first-order, especially since Labov (2010, 186) notes that Low Back Merger, the Northern Cities Shift, the Canadian Shift and others "all take place well below the level of social consciousness." In addition, not all imposed first-order correlations come to be socially meaningful (Johnstone & Kiesling 2008, 10). Nevertheless, attempts at describing the kinds of indexes that we could associate with the Reverse Vowel Shift at this stage, are provided below.

⁷⁶ Though the numbering of the indexes indicates linearity, this is not the case. The progression from first order to second and third order indexicality is a potential one, and does not always occur (Eckert 2008, 464; Johnstone & Kiesling 2008, 24).

The findings show that the younger speakers lead the reversal of the Older Vowel Shift: the younger the speaker, the lower and more retracted their short front vowels are. In this sense, the Reverse Vowel Shift indexes ‘youth’ on a first-order level, particularly youth of university-attending age (between 18 and 24). The shift is largely female-led, since women lower and retract the vowels more than men (except in the case of the DRESS vowel). However, the Reverse Vowel Shift does not index ‘female’ for two reasons. Firstly, men also lower and retract, which means that it is not a feature solely associated with women, as it is in Dublin English, for instance (Hickey 2013). Secondly, as Eckert (2008, 455) aptly states, “women (and men) are not saying ‘I’m a woman’ when they use a ‘female-led’ change, nor are they saying ‘I’m not a woman’ when they do not.”

The fact that DRESS is the only vowel that young women do not lower as much as young men indicates that the way in which this vowel is treated could index something beyond ‘youth’. Though I postulate that the higher variant used by young women is attempted maintenance of a perceived prestige variable, further work into this is required before solid claims about indexicality can be made. Tentatively, therefore, the raised DRESS variants used by the young women potentially indexes the desire to maintain something that they attribute to being a kind of ‘prestigious English-speaking South African’.

Though the low and retracted quality of the KIT, DRESS and TRAP sets in Hawai’i English are not related to a chain shift, there are similarities between it and SAfE. Specifically, the findings indicate that young Hawaiian men lower DRESS more than young women (Drager, Kirtley, Grama & Simpson 2013, 45). While the authors do not provide a reason for this distinction, they use it, along with proficiency in Pidgin, to prove that lowering and retraction, though similar to trends in California and Canada, are not the result of a chain shift. Instead, social work is being done, related to gender and the ability to speak Pidgin. In this sense it is likely there is more to be deduced from the gendered variation in DRESS lowering in SAfE.

The Reverse Vowel Shift in general is a move away from what was previously a characteristic way of sounding like an English-speaking South African (i.e. the raised short front vowels). Essentially, the shift is therefore a process of re-characterising what a South African sounds like, modelled in part, I argue, on international, particularly American, trends. If the Reverse Vowel Shift is a first-order index indicating ‘South African’, the fact that it

results from ‘contact’ with international Englishes could substantiate the claim that it also indexes ‘international awareness’ and ‘global participation’, perhaps on a second-order level.

Johnstone and Kiesling (2008, 5) argue for a ‘phenomenological’ approach to the study of indexicality, which is a consideration of the ways in which history and lived experiences give rise to the creation of meaning via language. Given the emergence of reverse shifting around the year 2000, it is possible that the lowering of the short front vowels is an attempt to articulate a ‘new’ middle class South African identity, indexing a disassociation with the ‘old’ South Africa. This idea is strengthened by the fact that, though the black speakers in this sample do not participate in the Reverse Vowel Shift, certain of its features (such as TRAP lowering) are common in their speech. These speakers therefore speak a crossover variety indicative of class, not ethnicity, which could also be construed as a move away from the notions of the ‘old’ South Africa. I cannot substantiate these claims regarding shifting indexicalities (or moves within an indexical field) without probing more deeply into ideologies of and attitudes towards English as it is used and perceived in Cape Town and perhaps other cities in South Africa.

A concept closely linked with indexicality is Agha’s (2003, 2007) notion of ‘enregisterment’, a process by which certain linguistic forms become linked with social meaning (Johnstone 2009, 159). Quoting Agha (2003, 231), Johnstone, Andrus and Danielson (2006, 79-80) further define the term as “the identification of a set of linguistic forms as a ‘linguistic repertoire differentiable within a language as a socially recognised register,’ which has ‘speaker status linked to a specific scheme of cultural values.’” Johnstone (2006) provides a good example of the enregisterment of various linguistic features, using the case of ‘Pittsburghese’. As mentioned previously, various Pittsburgh dialect features, such as MOUTH monophthongisation, have become indexed on the third-order level, and so have become inextricably linked to the notion of ‘being from Pittsburgh’. Johnstone (2006) argues that the t-shirts showcasing Pittsburghese dialect features is an example of and has contributed to the enregisterment of the dialect. In particular, Johnstone (2006, 168) notes,

(...) the shirts appeal to people who are able to hear Pittsburgh speech as different from other varieties and who link Pittsburgh speech not with working-class or incorrect speech as much as with authentic local identity.

In order for a dialect feature or a set of dialect features to be enregistered in this way, the indexicality linked to the features must be on a third-order level (Johnstone 2006, 168). In this sense, it is not yet possible to concretely posit any kind of enregisterment for the Reverse Vowel Shift in SAfE, since there are as yet no third-order indexes associated with it. Nevertheless, the way in which people perceive the shift, and how it is portrayed in the media can provide some insights.

A South African bank, *Capitec*, launched an advertisement in July 2015, set in a shopping centre, about a bank account for a fee of five Rand.⁷⁷ The advertisement follows a vox pop format, where three young South Africans are asked, “What can you get for five Rand nowadays?” The first respondent is a young man, who responds with, “Five Rand?” The second respondent is also a young man, who says, “Not a lot.” The third respondent is a young woman, who is offered a secondary question: “How about five Rand for a bank account?” She responds, “No, never! You can’t even get a chocolate for five Rand. How can you get a bank account for five Rand?”

The interest in this advertisement is the pronunciation of the words *rand* and *bank*, which both contain the TRAP vowel. Only two of the young people in the advert used these words, and via a quick auditory analysis, these two speakers (the first man and the woman) largely fit into the data I have analysed in the previous chapters. The young man is a black speaker of SAfE, and does not participate in the Reverse Vowel Shift. The young woman is a white speaker of SAfE, and realises TRAP as a very low, fairly retracted vowel (impressionistically, to [æ̘] or [ä]), which means that she also participates in the other features associated with the Reverse Vowel Shift. Interestingly, both the young man and the young woman produce the /æ/ in *rand* with significant creak, a feature pervasive in the speech of especially young white, middle class women (Bekker 2009, 432), but also to some extent among a number of young, black, middle class women (Wileman in progress).

Given the nature of the advertisement (vox pop, set in a shopping centre), and the smart clothing of the people in it, it is reasonable to assume that the intended message was to target young, fashionable, middle class South Africans. The advert does not strike me as spontaneous, since the way in which the young woman responds is clearly rehearsed. It is

⁷⁷ The advertisement was published on YouTube on 10 July 2015:
https://www.youtube.com/watch?v=-C80Xl_8psM

therefore also likely that the people in the advertisement were chosen as representative of the target market, in addition to which they were likely told to ‘sound natural’. The fact that both speakers creak when ‘sounding normal’, or at the very least, when being aware of being filmed and observed, indicates that it is a feature so common that it is not noticed. In addition, the use of a lowered and retracted TRAP (and by extension the Reverse Vowel Shift) is therefore seen as a natural way for a young, white, South African woman to speak. This in turn suggests that these features are (a) seen as appropriate to the context of marketing to a younger audience, and (b) a very common feature in the speech of particularly young white women, to the extent that it is not a feature likely to be perceived as incongruous.

To get a glimpse into how the Reverse Vowel Shift is perceived by listeners, and through this to ascertain levels of indexicality and enregisterment, I sent a link to the advertisement to first year students in the Linguistics Section at UCT and to friends and family. They were asked to watch the advertisement, and to comment on the young woman, with particular reference to where they thought she was from and anything else they could glean from the two sentences she utters. In total, 12 people responded, varied in age from 18 to 58.

In addition to the observations I gathered, there were two interesting comments on the advertisement placed on YouTube, which both suggest an aversion to the way the young woman speaks. One viewer expressly ‘hates’ the way she pronounces *rand*, and the other would like for her to ‘stop the twang’, as is shown in figure 6.9.



Figure 6.9: Responses on YouTube to an advertisement by *Capitec* Bank.

‘To twang’ or ‘twanging’ refers to what is perceived as ‘overcompensation’ when speaking English in an accent that is not ‘natural’ (also called a ‘rhaa rhaa’ (/ɪɑ: ɪɑ:/), alluding to the rolled nature of L1 English /r/ in SAfE). Black speakers of crossover SAfE are often said to ‘twang a lot’ or to ‘speak in a rhaa rhaa accent’, and are derisively referred to as *coconuts*

(i.e. white on the inside, brown on the outside).⁷⁸ Since ‘twanging’ is usually associated with ‘people trying to sound white’, it is interesting that the second comment suggests that the young woman twangs. In this sense it would seem that the way she speaks is considered fake or overly posh, or that she perhaps ‘tries too hard’. Trying too hard, and thereby producing a very low, retracted TRAP suggests a level of prestige associated with the Reverse Vowel Shift.

The respondents to my call for opinions did not express negative attitudes towards the young woman’s accent, though the notion of ‘twanging’ is echoed in some of the responses, perhaps more linked to ‘poshness’ rather than a ‘rhaa rhaa’ accent. Some respondents commented that her accent sounds ‘posh’ or ‘snobbish’, or is indicative of someone in the upper social class, likely to have gone to a private school. Of the 15 respondents, 10 made such comments. It would therefore appear that short front vowel lowering is associated with ‘poshness’ and with those on the higher end of the socioeconomic scale. Interestingly, the notion of ‘being posh’ is also linked to short front vowel lowering in Dublin (Hickey, 2013).

With regards the association of her accent with a particular place, the responses were divided. Of the 15 responses, seven indicated that she sounds like someone from Cape Town, while six stated she was from Johannesburg, and one from Durban. A single respondent noted that the accent was “fairly neutral”. In some cases, the respondents provided a second option for where the young woman was from, and these included Cape Town (n = 1), Johannesburg (n = 4) and Durban (n = 2). The Reverse Vowel Shift is therefore associated with big South African cities, as expected, most strongly Cape Town and Johannesburg.

Important to note is that most respondents did not mention the way in which the young woman pronounced *rand* or *bank*, which indicates that they are not aware of the features upon which they based their assumptions. Only one respondent, aged 29, specifically noted the pronunciation of *rand*, and linked it with being South African: “The way she says “rand” is a dead give away that she is most likely South African.” This respondent then goes on to guess that the speaker is from Johannesburg or Durban. I queried her response, asking her exactly what it was about the pronunciation of *rand* that sounds South African, to which she replied:

⁷⁸ Thanks to Ndumiso Madubela for this formulation of ‘twang’.

Lol not sure, it just sounds typically South African. If I were to try and explain it I would say "rund" instead of "raaaand". Like a deep "a" vowel. (...) It's almost as if she drops the sound of her vowels. And I think it's that deep, rounded way of speaking that sounds typically South African to me.

The use of 'deep' and 'drops the sound of her vowels' alludes to the lowering associated with the Reverse Vowel Shift. Interestingly, short front vowel lowering is a relatively new phenomenon, and yet it sounds 'typically South African' to this respondent. Though she is the only one to observe the lowering of /æ/, albeit not overtly, the association of it with 'South Africanness' indicates that the feature has become indexical of place, at least to this respondent. In general, however, it seems reasonable to assume that the correlations drawn between accent and place/social standing do not involve overt awareness of specific dialectal features. In this way it seems unlikely that there are any second- or third-order indexes that can be attributed to the Reverse Vowel Shift, and in turn, that enregisterment is not (yet) underway. Of course, this requires more thorough investigation.

6.4 SUMMARY OF OBJECTIVES AND FINDINGS

This study on SAfE as spoken by a subset of Capetonians set out to answer three broad questions, specifically related to the emergence of a new vowel chain shift, potential merger between TRAP and STRUT, deracialisation and participation in international trends. Via the analysis of 53 speakers of SAfE using the FAVE suite and various statistical methods, each of these questions has been answered. I summarise the findings below, following the outline of the study provided in Chapter 1, section 1.1.

1. To what extent is the Reverse Short Front Vowel Shift present in SAfE?

- a. How far has the change progressed?

The analysis presented in Chapter 4 suggests that the Reverse Vowel Shift is well under way in SAfE. This shift is a combination of push and pull chains. The push chain concerning FOOT and KIT is initiated by the fronting of FOOT, causing the subsequent lowering of KIT to avoid merger. A pull chain is evidenced and initiated by the lowering of TRAP, which not only undoes the vowel raising associated with the Older Vowel Shift, but also reverses it by the extreme lowering. In addition to this lowering, TRAP retracts towards [a], which is an innovative feature in that the retraction does not undo any process associated with the Older Vowel Shift. In retracting, TRAP pushes STRUT away in a push chain. The evidence furthermore supports the lowering of DRESS in response to lowered TRAP. However, DRESS has not yet lowered beyond [ɛ], the original position before the Older Vowel Shift. Hence, DRESS lowering undoes the Older Vowel Shift, but does not reverse it. KIT and its two major allophones, (1) around velars, after /h/, word initially (KIT2) and (2) elsewhere, show evidence of further retraction as predicted by Lanham (1965) as a feature of the Older Vowel Shift. Since both allophones retract, the overall split is maintained. In addition to this retraction, DRESS also retracts, perhaps as a generalisation of the backing of KIT and TRAP.

- b. Who participates in or leads the change?

The shift is led by young, middle class speakers. The younger the speaker, the more retracted and lowered their short front vowels are. In most cases, young women lead the lowering and retracting, producing the most shifted realisations. This is not the case, however, for the DRESS vowel, where young men lower the vowel more than the young women do. In this case, it appears that the women perceive a higher DRESS vowel as a prestige variable, and have as a result resisted the lowering initially. There are thus two DRESS variables in the shift:

the lower, male version [ɛ̃], and the higher, female version [ɛ̄], though it is important to note that both versions are lower than DRESS associated with the Older Vowel Shift (i.e. [e] or [ɪ]).

- c. To what extent do black speakers participate in this newer vowel shift, and thus what claims can be made about convergence and deracialisation?

Though there are many similarities between the two ethnicities, the overall conclusion is that the black speakers do not fully participate in the Reverse Vowel Shift. Though the black women in the sample participate in the KIT split, the secondary backing of DRESS_1 and TRAP retraction, the participation is not yet consistent enough to regard the Reverse Vowel Shift as a regular feature of their speech. This is supported by the fact that they do not lower or retract DRESS when compared to the white speakers. Moreover, the black men in the sample are consistently different from the white speakers, retaining variables that are similar to the descriptions of traditional Black South African English, such as the lack of a KIT split; or acrolectal Black South African English, such as the realisation of TRAP as [æ].

That the black women show the most similarity to the young white speakers is indicative of partial convergence. The similarities between the white and black speakers, though not categorical, show changes to the ethnic associations once made for English. Young, middle class, black, particularly female South Africans speak a crossover variety, in the sense that it is no longer similar to traditional Black South African English: it is crossing over into a new linguistic space formerly associated with white speakers. However, the Reverse Vowel Shift may be considered a new divergence from older white norms and possibly newer, black ‘crossover’ norms. Mesthrie (2015, personal communication) notes that such new divergence via the Reverse Vowel Shift would be quite a radical move in the wake of recent crossover speech. He asks: “have young, white, middle class women unconsciously innovated and stabilised the Reverse Vowel Shift, thereby maintaining a distance from other converging groups?” This would raise the spectre of the sociologist’s notion of ‘elite’ flight or ‘white flight’ if, just when some black speakers appeared to match the white speakers in their crossover variety, white speakers innovate further. Such a strong claim with its significant social and political undertones would have to be probed more carefully, preferably in Johannesburg, the largest and earliest centre of innovation. It would also require a closer study of other individuals (e.g. coloured and Indian females) in private and model C schools.

2. Are /æ/ and /ʌ/ in a process of vowel merger?
 - a. If so, how far along is the change, if not, what leads to the overlap?

The comprehensive analysis in Chapter 5 shows that TRAP and STRUT do not merge. Instead, STRUT is the latest instalment in the story of the Reverse Vowel Shift, raising and backing as a result of lowering and retracting TRAP. The overlap suggesting merger in some tokens is temporary, especially since push chains often result in a high degree of token overlap (Labov 2010, 144).

3. What patterns can be found in the changing vowels of SAfE, and to what extent do they overlap with patterns found in other Englishes?

- a. If there are global patterns, can these be attributed to internal and external factors?

The short front vowel lowering present in Capetonian SAfE is found in Englishes throughout the world. As mentioned before, Canada, California, southeast England, Ireland and Australia also lower these vowels in a chain shift. In addition, trends of short front vowel lowering are also emerging in varieties of American Englishes associated with the Northern Cities and the Southern Shift, both of which are generally associated with vowel raising. A global trend, therefore, is the lowering of KIT, DRESS and TRAP, often paired with retraction. Where the varieties vary is in the impetus for the shifts. The pivot points of chain shifts as described by Labov (1994) are Low Back Merger or the status of TRAP (i.e. tense/lax allophones and if it raises or not). Low Back Merger is the instigator in California and Canada, while the lowering and retraction of TRAP is the instigator in South Africa (in conjunction with FOOT), Australia, Ireland and southeast England.

The Reverse Vowel shift in particular is a combination of internal and external factors. The endogenous fronting of FOOT is causing KIT to lower, in addition to which the lowering of TRAP, leading to STRUT raising and backing, can potentially be seen as a change due to colonial lag (though this notion requires historical re-investigation) sped up by virtual and physical contact with TRAP lowering varieties of British and American Englishes.

6.5 SIGNIFICANCE AND IMPACT OF THE STUDY

Though there has been much research on varieties of SAfE spoken by white members of the population, this is the first to thoroughly document the new vowel shift noticed by Bekker (2009) and Mesthrie (2012a). Regarding the short vowels in general, this thesis complements the word list style data presented in Bekker (2009). Furthermore, this thesis highlights the participation of South Africans in global trends, which has hitherto not been described in much detail, barring the comparisons of Bekker (2009). It is hoped that the findings presented regarding the Reverse Vowel Shift can act as a basis for further research into this phenomenon and indeed SAfE in general.

The comparisons between black and white speakers of SAfE partially support the findings of convergence and deracialisation of particularly Mesthrie (2010), and in this way contributes to the body of work surrounding social change in post-segregation South Africa.

Regarding method, it is the first acoustic study of SAfE to utilise the FAVE suite. Furthermore, the main statistical methods employed are fairly new to sociolinguistics in South Africa, and in this sense it is hoped that the analysis presented in this thesis will provide statistical methodology for further large-scale studies in South Africa.

6.6 FUTURE ENDEAVOURS

In addition to the confirmation of the emerging Reverse Vowel Shift in SAfE, this study has brought to light various avenues of research into SAfE that have as yet either been neglected, or require further investigation. This thesis has, during the course of the analyses and discussions, raised several questions that it is unable to answer. Specifically, future research should endeavour to provide more and precise information on the following topics:

1. The status of BATH in SAfE, specifically in Cape Town.
2. The long vowel system of SAfE, specifically related to an apparent anti-clockwise chain shift, and in addition, if this has any effect on short front vowel lowering. If one can show that the movement of the long vowels, particularly GOOSE and THOUGHT, leads to an overcrowding of the high front vowel space, it could be a further internal reason for the lowering of KIT and DRESS in addition to a pull chain initiated by TRAP, which is linked to the rotation principle (Hickey 2014).

3. The characteristics of LOT and how it relates to the changes observed as the Reverse Vowel Shift.
4. The particulars of the DRESS vowel, to confirm whether men lead the change, or whether there are indeed two gendered variants of the vowel.
5. The effect of style. Given the focus on casual style, a natural follow-up would be to test the hypotheses of this study using word list or formal speech data, to see if style shifting affects the use of the Reverse Vowel Shift in Cape Town.
6. The overlap between KIT and DRESS. This should be investigated further, using minimal pairs and commutation testing (Labov 2004, 356). In addition, a more in depth diachronic study of when exactly KIT and DRESS each began to lower is necessary. This will entail work with South Africa's sound archives.
7. The Reverse Vowel Shift in other geographical areas. The study should be expanded to include other big cities, such as Johannesburg and Durban in addition to smaller centres, in order to ascertain whether or not the Reverse Vowel Shift is a big-city phenomenon. Replicating this study particularly in Johannesburg, with casual style data, would be highly valuable regarding insights into convergence. Middle class black South Africans are more prominent in Johannesburg, and a study comparing middle class black and white speakers would allow us to gauge more accurately whether or not there are shared norms.
8. The indexicality and potential enregisterment of the Reverse Vowel Shift. Further qualitative research, such as the matched-guise tests employed by Johnstone and Kiesling (2008), would provide greater insight into the social correlations of the changes observed.

In addition to these areas of expansion, there are shortcomings in the sample that should be ironed out in future research, particularly regarding sample size. A larger number of black speakers would aid more accurate descriptions of post-segregation change. Generally speaking, more speakers between 35 and 95 years of age would provide scope for thorough apparent time analysis, which in turn would likely illuminate some of the questions raised, particularly with reference to the adoption and stabilisation of the Reverse Vowel Shift in Cape Town.

6.7 CONCLUSION

This thesis, via new analytical methods, has shown that the incipient Reverse Vowel Shift noticed by Bekker (2009) and Mesthrie (2012a) is now a regular feature of middle class English in Cape Town. In addition, it has shown that there is a large degree of global participation in the lowering of the short front vowels, coupled with innovations surrounding vowel retraction. Partial convergence and deracialisation of English in Cape Town is a further finding, showing the impact of social change on the face of language use in South Africa.

APPENDICES

APPENDIX A

Chapter 3: Campus-wide email calling for volunteers, approved by the Humanities Research Ethics Committee.

Dear Student,

I am doing research based in Cape Town for my doctorate in Linguistics. The collection of language data involves talking to people who have lived in Cape Town for the majority of their lives, and who speak English fluently. If you are from Cape Town and are between 18 and 30, I would love to talk to you at your convenience. Please send me an email and we can organise a time to suit you.

What is required from the participant: Participation in a recorded conversation, on upper campus, about their background, daily life and experiences in Cape Town. Participants will also be asked to read a list of words. No information from third parties will be collected without their consent.

Risks: None

Benefits: Contributing to sociolinguistic research on varieties of and changes to English as spoken in South Africa.

Costs: None to the participant.

Each volunteer signs a consent form prior to the interview, which assures anonymity and confidentiality. Furthermore it states that the collected data will be used for research purposes and in aggregate form only, so that the participant cannot be identified. A volunteer can withdraw their participation at any point before, during or after the interview.

To volunteer, or if you have questions, please email alida.chevalier@gmail.com or phone me at [021 650 5907](tel:0216505907).

If you would like to know about the project as a whole (this doctorate forms part of a broader project on Language in South Africa, funded by the NRF), please email Prof Rajend Mesthrie, the SARCHI Chair of the project and my supervisor: rajend.mesthrie@uct.ac.za.

I look forward to meeting you,

Alida

Chapter 3: consent form signed by each participant.



University of Cape Town
Linguistics Section
School of African and Gender Studies, Anthropology and Linguistics
CONSENT FORM FOR RESEARCH ON HUMAN SUBJECTS

1. Topic of research project:

The Reverse Vowel Shift in South African English

2. Names of principal researcher(s):

Alida Chevalier

3. Department or research group address:

Linguistics Section
Room 14 Arts Block
University of Cape Town
7701 Rondebosch

4. Contact details:

☎: 021 650 5907 ✉: alida.chevalier@gmail.com

5. Name of participant:

6. Nature of the research:

Quantitative, sociolinguistic, acoustic, ethnographic.

7. Participant's involvement:

What is required from the participant: Participation in a recorded interview about their background, daily life and experiences in Cape Town. Participants will also be asked to read a list of words and a reading passage. No information from third parties will be collected without their consent.

Risks: None

Benefits: Contributing to sociolinguistic research on varieties of and changes to English as spoken in South Africa.

Costs: None to the participant.

- I agree to participate in this research project.
- I have read the consent form and the information it contains and had the opportunity to ask questions about them.
- I agree to my responses being used for education and research on condition that my privacy is respected, subject to the following:
 - I understand that my personal details will be used in aggregate form only, so that I will not be personally identifiable.
- I understand that I am under no obligation to take part in this project.
- I understand that I have the right to withdraw from this project at any stage.

Signature of participant _____

Date: _____

APPENDIX B

Chapter 3: word list and minimal pairs test, presented to each participant.

WORD LIST		MINIMAL PAIRS TEST
KIT	BID	CAT
DRESS	BED	CUT
TRAP	BAD	
LOT	BOD (BODY)	LUCK
STRUT	BUD (BUDDY)	LACK
FOOT	BUD (BUDDHA)	
BATH	BARD	MONEY
CLOTH	BIRD	MANY
NURSE	BEAD	
FLEECE	BADE	MATCH
FACE	BORED	MUCH
PALM	BODE	
THOUGHT	BOED	PUNT
GOAT	BIDE	PANT
GOOSE	BOYD	
PRICE	BOWED	BAT
CHOICE	BEARD	BUT
MOUTH	BARED	
NEAR		BACK
SQUARE		BUCK
START		
NORTH		PUCK
FORCE		PACK
CURE SIT		
GET		PAT
CAT		PUTT
COT		
BUT		CUP
PUT		CAP
PARK		
POT		RANT
HEARSE		RUNT
EAT		
ATE		
LUCKY		
CAUGHT		
ODE		
HOOP		
DICE		
BOY		
SHOUT		
FEAR		
FAIR		
BATTER		
LOVE		
BUTTER		
MONEY		

APPENDIX C

Chapter 6: Distance measures between TRAP and STRUT, casual style results for all speakers.

Speaker	Mean	TRAP	STRUT	Euclidean Distance	t-score	p Welch	Pillai Score	p
AaronA	F1	802	720	513	6.903	0.000	0.585	0.000
	F2	1981	1475		17.653	0.000		
AlanL	F1	834	725	341	12.33	0.000	0.710	0.000
	F2	1701	1379		18.017	0.000		
BobbyC	F1	789	808	217	-1.227	0.223	0.442	0.000
	F2	1805	1588		7.025	0.000		
BradP	F1	834	774	264	5.695	0.000	0.479	0.000
	F2	1715	1460		11.998	0.000		
BrendaF	F1	725	706	393	0.744	0.461	0.706	0.000
	F2	1866	1473		11.776	0.000		
CaitlynS	F1	863	739	218	10.023	0.000	0.668	0.000
	F2	1575	1396		10.137	0.000		
ChloëV	F1	830	746	139	4.518	0.000	0.335	0.000
	F2	1597	1486		5.383	0.000		
CindyH	F1	837	755	395	6.428	0.000	0.631	0.000
	F2	1769	1383		19.125	0.000		
CraigS	F1	780	784	375	-0.556	0.579	0.580	0.000
	F2	1865	1490		20.4	0.000		
EricaV	F1	867	725	174	7.451	0.000	0.382	0.000
	F2	1549	1450		3.60	0.001		
GaryL	F1	870	753	294	7.622	0.000	0.471	0.000
	F2	1664	1395		11.302	0.000		
GeoffP	F1	810	740	275	6.233	0.000	0.539	0.000
	F2	1763	1497		12.751	0.000		
GlenD	F1	783	721	364	6.861	0.000	0.576	0.000
	F2	1832	1473		17.375	0.000		
HayleyS	F1	827	764	257	5.090	0.000	0.416	0.000
	F2	1703	1454		9.879	0.000		
IlseP	F1	840	737	131	9.082	0.000	0.278	0.000
	F2	1567	1486		3.801	0.000		
JaneK	F1	838	731	186	11.139	0.000	0.453	0.000
	F2	1669	1517		11.702	0.000		
JanetH	F1	842	762	236	6.733	0.000	0.626	0.000
	F2	1664	1441		11.485	0.000		
JennyL	F1	825	723	207	6.482	0.000	0.466	0.000
	F2	1611	1429		8.690	0.000		
JerryM	F1	841	762	300	6.462	0.000	0.633	0.000
	F2	1650	1360		15.258	0.000		
JimH	F1	770	767	356	0.117	0.908	0.710	0.000
	F2	1876	1520		9.483	0.000		
KaleyW	F1	815	728	241	4.806	0.000	0.431	0.000
	F2	1685	1460		9.173	0.000		
KaraL	F1	809	749	157	4.071	0.000	0.302	0.000
	F2	1685	1540		7.036	0.000		

KayC	F1	830	764	392	3.489	0.001	0.636	0.000
	F2	1796	1409		13.033	0.000		
LarryR	F1	807	706	186	5.233	0.000	0.359	0.000
	F2	1664	1508		3.583	0.001		
LauraE	F1	899	785	197	5.138	0.000	0.452	0.000
	F2	1643	1475		8.723	0.000		
LeeL	F1	841	770	181	5.664	0.000	0.456	0.000
	F2	1660	1493		9.778	0.000		
LexiK	F1	860	766	233	4.467	0.000	0.447	0.000
	F2	1678	1464		6.783	0.000		
LiamH	F1	851	766	336	5.780	0.000	0.745	0.000
	F2	1798	1473		16.267	0.000		
MarilynK	F1	779	788	309	-0.73	0.4665	0.582	0.000
	F2	1776	1467		17.707	0.000		
MikeF	F1	793	761	332	2.055	0.042	0.669	0.000
	F2	1815	1484		12.916	0.000		
MilesG	F1	831	756	417	8.790	0.000	0.514	0.000
	F2	1832	1422		17.092	0.000		
MindyG	F1	846	797	149	3.556	0.001	0.321	0.000
	F2	1663	1521		8.298	0.000		
PaddyF	F1	797	770	337	1.900	0.059	0.528	0.000
	F2	1792	1456		15.488	0.000		
PatrickM	F1	869	770	251	10.153	0.000	0.665	0.000
	F2	1664	1433		14.261	0.000		
PerryF	F1	747	727	285	0.704	0.490	0.424	0.000
	F2	1824	1540		5.624	0.000		
RachelJ	F1	844	757	233	4.990	0.000	0.568	0.000
	F2	1641	1424		7.603	0.000		
RichS	F1	797	717	342	7.292	0.000	0.447	0.000
	F2	1790	1457		10.891	0.000		
RuthW	F1	840	749	201	5.182	0.000	0.465	0.000
	F2	1599	1420		8.758	0.000		
SamuelB	F1	834	754	289	7.579	0.000	0.610	0.000
	F2	1715	1437		14.176	0.000		
SarahN	F1	828	724	259	5.572	0.000	0.407	0.000
	F2	1704	1467		6.612	0.000		
TamiH	F1	785	707	216	3.877	0.000	0.531	0.000
	F2	1635	1434		7.045	0.000		
TedR	F1	798	707	308	8.849	0.000	0.695	0.000
	F2	1763	1434		26.614	0.000		
TodK	F1	808	770	260	3.90	0.000	0.491	0.000
	F2	1722	1465		14.349	0.000		
ValV	F1	745	815	434	-6.973	0.000	0.648	0.000
	F2	1904	1475		25.167	0.000		

WORKS CITED

- Adendorff, R., & Savini-Beck, M. (1993). The teaching of English vowels and consonants in the new South Africa. *Journal of Language Teaching*, 27, 232-248.
- Agha, A. (2003). The Social Life of Cultural Value. *Language and Communication*, 23, 231-273.
- Agha, A. (2007). *Language and Social Relations*. New York: Cambridge University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7,. URL: <http://CRAN.R-project.org/package=lme4>.
- Bauer, L. (1979). The second great vowel shift. *Journal of the International Phonetic Association*, 9, 57-66.
- Bauer, L. (1992). The second great vowel shift revisited. *English World-Wide*, 13, 253-268.
- Bauer, L. (1994). *Watching English Change*. London: Longman.
- Bekker, I. (2007). Fronted /s/ in General White South African English. *Language*, 38 (1), 46-74.
- Bekker, I. (2009). The Vowels of South African English. PhD Thesis: North-West University.
- Bekker, I. (2012). South African English as a late 19th-century extraterritorial variety. *English World-Wide*, 33 (2), 127-146.
- Bekker, I. (2013a). The Formation of South African English. *English Today*, 29 (1), 3-9.
- Bekker, I. (2013b). Nursing the Cure: a phonetic analysis of /oə/ in South African English. *Stellenbosch Papers in Linguistics Plus*, 42, 1-39.
- Bekker, I. (2014a). The KIT-Split in South African English: A critical review. *Southern African Linguistics and Applied Language Studies*, 32 (1), 113-131.
- Bekker, I. (2014b). The weak vowels of South African English: A critical review and comparative acoustic analysis. *Southern African Linguistics and Applied Language Studies*, 32 (1), 133-147.
- Bekker, I., & Eley, G. (2007). An acoustic analysis of White South African English (WSAfE) monophthongs. *Southern African Linguistics and Applied Language Studies*, 25 (1), 107-114.
- Bernard, J. (1967). Some Measurements of Some Sounds of Australian English. PhD Thesis: Sydney University.

- Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer [Computer program]. Version 5.3.63*. Retrieved January 25, 2014 from <http://www.praat.org/>
- Bowerman, S. (2004). White South African English: phonology. In E. W. Schneider, K. Burrige, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English*. (Vol. 1: Phonology, pp. 931-942). Berlin: Mouton de Gruyter.
- Bradley, D. (1989). Regional dialects in Australian English phonology. In P. Collins, & D. Blair (Eds.), *Australian English: The Language of a New Society* (pp. 260-270). St Lucia: University of Queensland Press.
- Branford, W. (1994). English in South Africa. In R. Burchfield (Ed.), *The Cambridge History of the English Language, vol. 5: English in Britain and overseas - origins and development*. (pp. 182-429). Cambridge: Cambridge University Press.
- Breckwoldt, G. (1961). Some aspects of the phonetics of vowels in South African English monophthongs. *Maitre Phonétique*, 115, 5-12.
- Brown, J. (2012). Focusing and Diffusion in 'Cape Flats English': a Sociophonetic Study of Three Vowels. MA Thesis: University of Cape Town.
- Buschfeld, S., Hoffmann, T., Huber, M., & Kautzsch, A. (Eds.). (2014). *The Evolution of Englishes: the Dynamic Model and beyond*. Amsterdam/Philadelphia: John Benjamins.
- Chambers, J. K. (1998). TV Makes People Sound the Same. In L. Bauer, & P. Trudgill (Eds.), *Language Myths* (pp. 123-131). London: Penguin Books.
- Chambers, J. K. (2004). 'Canadian Dainty': the rise and decline of Briticisms in Canada. In R. Hickey (Ed.), *Legacies of Colonial English: studies in transported dialects* (pp. 224-241). Cambridge: Cambridge University Press.
- Chevalier, A. (2011). Social class differentiation in South African Indian English: a sociophonetic study of three vowel variables. MA Thesis: University of Cape Town.
- Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, 7 (2), 209-228.
- Coetzee-Van Rooy, S. (2014). The identity issue in bi- and multilingual repertoires in South Africa: Implications for Schneider's Dynamic Model. In S. Buschfeld, T. Hoffmann, M. Huber, & A. Kautzsch (Eds.), *The Evolution of Englishes: the Dynamic Model and beyond*. (pp. 39-57). Amsterdam/Philadelphia: John Benjamins.
- Cox, F. (1999). Vowel change in Australian English. *Phonetica*, 56, 1-27.
- Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, 26, 147-179.

- Cox, F., & Palethorpe, S. (2001). The changing face of Australian English vowels. In D. Blair, & P. Collins (Eds.), *English in Australia: Varieties of English Around the World*, vol 26. Amsterdam: John Benjamins.
- Cox, F., & Palethorpe, S. (2004). The Border Effect: vowel differences across the NSW/Victorian border. In C. Moskovsky (Ed.), *Proceedings of the 2003 Conference of the Australian Linguistics Society*.
- Cox, F., & Palethorpe, S. (2007). Illustrations of the IPA: Australian English. *Journal of the International Phonetic Association*, 27, 341-350.
- Cox, F., & Palethorpe, S. (2008). Reversal of Short Front Vowel Raising in Australian English. *9th Annual Conference of the International Speech Communication Association 2008 (INTERSPEECH 2008)* (pp. 342-345). International Speech Communication Association.
- Cox, F., Palethorpe, S., & Tskuada, K. (2004). *A Century of Accent Change in Australian English*. Sydney: Paper presented at the 10th Australian International SST Conference.
- Da Silva, A. B. (2007). South African English: a sociolinguistic investigation of an emerging variety. PhD Thesis: University of the Witwatersrand.
- De Klerk, V. (2000). Language shift in Grahamstown: a case study of selected Xhosa speakers. *International Journal of the Sociology of Language*, 146, 87-110.
- Deepayan, S. (2008). *Lattice: Multivariate Data Visualization with R*. New York: Springer-Verlag.
- Delbridge, S. A. (2006). Aspiration in South African Indian English. MA Thesis: University of Cape Town.
- Den Besten, H. (1989). From Khoekhoe foreigner talk via Hottentot Dutch to Afrikaans: the creation of a novel grammar. In M. Putz, & R. Dirven (Eds.), *Wheels within wheels*. (pp. 207-254). Frankfurt: Peter Lang.
- Di Paolo, M., & Faber, A. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change*, 2, 155-204.
- Di Paolo, M., & Yaeger-Dror, M. (2011). Field Methods: gathering data, creating a corpus, and reporting your work. In M. Di Paolo, & M. Yaeger-Dror (Eds.), *Sociophonetics: a Student's Guide* (pp. 7-23). Oxon: Routledge.
- Dodsworth, R., & Kohn, M. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change*, 12, 221-245.

- Drager, K., Kirtley, M. J., Grama, J., & Simpson, S. (2013). Language Variation and Change in Hawai'i English: KIT, DRESS, and TRAP. *University of Pennsylvania Working Papers in Linguistics*, 19 (2), 41-50.
- Du Plessis, D., & Bekker, I. (2014). To err is human: The case for neorhoticity in White South African English. *Language Matter*, 45 (1), 23-39.
- Durian, D. (2012). A new perspective on vowel variation across the 19th and 20th centuries in Columbus, OH. PhD Thesis: The Ohio State University.
- Eckert, P. (2004). *Vowel Shifts in Northern California and the Detroit Suburbs: Northern California Vowels*. From <http://www.stanford.edu/~eckert/vowels/html>.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12 (4), 453-476.
- Evanini, K. (2009). The Permeability of Dialect Boundaries: a case study of the region surrounding Erie, Pennsylvania. PhD Thesis: University of Pennsylvania.
- Fabricius, A. H. (2007). Variation and change in the TRAP and STRUT vowels of RP: a real time comparison on five acoustic data sets. *Journal of the International Phonetic Association*, 37 (3), 293-320.
- Feagin, C. (1986). *Variation and Change in Alabama English: A Sociolinguistic Study of the White Community*. Washington, DC: Georgetown University Press.
- Finn, P. (2004). Cape Flats English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English*. (Vol. 1: Phonology, pp. 964-984). Berlin: Mouton de Gruyter.
- Fisiak, J. (1988). Karl Luick and historical English phonology. In D. Kastovsky, G. Bauer, & J. Fisiak (Eds.), *Luick Revisited: papers read at the Luick Symposium at Schloß Liechtenstein 15, 1985* (pp. 15-30). Tübingen: Gunter Narr Verlag.
- Gimson, A. C. (1962). *An Introduction to the Pronunciation of English*. London: Edward Arnold.
- Gordon, E., & Maclagan, M. (2001). 'Capturing a Sound Change': A Real Time Study Over 15 Years of the NEAR/SQUARE Diphthong Merger in New Zealand English. *Australian Journal of Linguistics*, 21 (2), 215-238.
- Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (Eds.). (2004). *New Zealand English: Its origin and evolution*. Cambridge: Cambridge University Press.

- Gordon, M. J. (2002). Investigating Chain Shifts and Mergers. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 244-266). Oxford/Massachusetts, USA/UK: Blackwell.
- Gordon, M. J. (2004). The West and Midwest: Phonology. In E. W. Schneider, K. Burrige, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of the Varieties of English*. (Vol. 1: Phonology, pp. 388-350). Berlin: Mouton de Gruyter.
- Gordon, M. J. (2011). Methodological and Theoretical Issues in the Study of Chain Shifting. *Language and Linguistics Compass*, 5 (11), 784-794.
- Hall-Lew, L. (2010). Improved representation of variance in measure of vowel merger. *Proceedings of Meetings on Acoustics 9(060002)*. doi:10.1121/1.3460625.
- Hartmann, D., & Zerbian, S. (2009). Rhoticity in Black South African English - a sociolinguistic study. *Southern African Linguistics and Applied Language Studies*, 27 (2), 135-148.
- Haudricourt, A. G., & Juilland, A. G. (1949). *Essai pour une histoire structurale du phonetisme francais*. Paris: C. Klincksieck.
- Hay, J., Drager, K., & Warren, P. (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, J., Maclagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press.
- Herold, R. (1990). Mechanisms of merger: The implementation and distribution of the low back merger in eastern Pennsylvania. PhD Thesis: University of Pennsylvania.
- Hickey, R. (1999). Dublin English: current changes and their motivation. In P. Foulkes, & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 265-281). London: Arnold.
- Hickey, R. (2005). *Dublin English: Evolution and Change*. Amsterdam: John Benjamins.
- Hickey, R. (2007). *Irish English: History and present-day forms*. Cambridge: Cambridge University Press.
- Hickey, R. (2013). *Variation and Change in Dublin English*. Retrieved July 22, 2014 from <https://www.uni-due.de/VCDE/>
- Hock, H. H. (1991). *Principles of Historical Linguistics*. (Second ed.). Berlin/New York: Mouton de Gruyter.

- Holmes, J., & Bell, A. (1990). Attitudes, varieties, discourse: an introduction to the sociolinguistics of New Zealand English. In A. Bell, & J. Holmes (Eds.), *New Zealand Ways of Speaking English* (pp. 1-20). Clevedon: Multilingual Matters.
- Hooper, A. (1945). A preliminary report of an investigation of spoken English in South Africa. *South African Journal of Science*, 41, 476-484.
- Hooper, A. (1951). English in South Africa. In E. Partridge, & J. Clark (Eds.), *British and American English since 1900*. (pp. 80-84). London: Andrew Dakers.
- Hopwood, D. (1928). *South African English Pronunciation*. Cape Town: Juta and Co.
- Horvath, B. (2004). Australian English: phonology. In E. W. Schneider, K. Burrige, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (Vol. 1: Phonology, pp. 625-644). Berlin: Mouton de Gruyter.
- Hothorn, T., & Zeileis, A. (2014). partykit: A Modular Toolkit for Recursive Partitioning in R. *Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universitaet Innsbruck*.
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival Ensembles. *Biostatistics*, 7 (3), 355-373.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15 (3), 651-674.
- Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2015). Package 'party'. <https://cran.r-project.org/web/packages/party/party.pdf>.
- Hundleby, C. E. (1964). Xhosa-English pronunciation in the South-East Cape. PhD Thesis: Rhodes University.
- Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics*, 39, 683-693.
- Jakobson, R. (1971). *Selected Writings III*. The Hague: Mouton.
- Jespersen, O. (1949). *A Modern English Grammar on Historical Principles. Part I: Sounds and Spellings*. London: George Allen & Unwin.
- Johnstone, B. (2009). Pittsburghese Shirts: commodification and the enregisterment of an urban dialect. *American Speech*, 84 (2), 157-175.
- Johnstone, B., & Kiesling, S. F. (2008). Indexicality and experience: exploring the meanings of /aw/-miniphthongisation in Pittsburgh. *Journal of Sociolinguistics*, 12 (1), 5-33.

- Johnstone, B., Andrus, J., & Danielson, A. E. (2006). Mobility, Indexicality and the Enregisterment of "Pittsburghese". *Journal of English Linguistics*, 34 (2), 77-104.
- Kamata, M. (2008). An Acoustic Sociophonetic Study of Three London Vowels. PhD Thesis: The University of Leeds.
- Kaplan, M., & Robertson, M. (1991). Introduction. In M. Kaplan, & M. Robertson (Eds.), *Founders and Followers: Johannesburg Jewry, 1887-1915*. (pp. 10-16). Cape Town: Vlaeberg.
- Keller, R. (1994). *On Language Change: the invisible hand in language*. London: Routledge.
- Kendall, T., & Thomas, E. R. (2012). vowels: Vowel Manipulation, Normalization, and Plotting. R package version 1.2. URL: <http://CRAN.R-project.org/package=vowels>.
- Kennedy, R., & Grama, J. (2012). Chain shifting and centralisation in California vowels: an acoustic analysis. *American Speech*, 87 (1, Spring), 39-55.
- Kuryłowicz, J. (1965). A remark on the Great Vowel Shift. *Word*, 21, 183-187.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-11., <http://CRAN.R-project.org/package=lmerTest> .
- La Page, R. B., & Tabouret-Keller, A. (1985). *Acts of Identity: creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Labov, W. (1979). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, 57, 267-309.
- Labov, W. (1991). The Three Dialects of English. In P. Eckert (Ed.), *New Ways of Analysing Sound Change* (pp. 1-44). San Diego: Academic Press.
- Labov, W. (1994). *Principles of Linguistic Change Volume 1: Internal Factors*. Cambridge, Oxford: Blackwell.
- Labov, W. (2010). *Principles of Linguistic Change Volume 3: Cognitive and Cultural Factors*. Malden: Wiley-Blackwell.
- Labov, W., & Wald, B. (1969). *General Principles of Vowel Shifting*. Paper given at the winter meeting of the LSA, San Fransisco.
- Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. New York: Mouton de Gruyter.
- Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language*, 89 (1), 30-65.

- Labov, W., Yaeger, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress*. Philadelphia: US Regional Survey.
- Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Boston: Thomson Wadsworth.
- Langstrof, C. (2006). Acoustic evidence for a push-chain shift in the Intermediate Period of New Zealand English. *Language Variation and Change*, 18, 141-164.
- Lanham, L. W. (1965). South African English pronunciation 2065 A.D. *Taalfasette*, 1 (1), 89-95.
- Lanham, L. W. (1967). *The Pronunciation of South African English: A Phonetic-Phonemic Introduction*. Cape Town: A.A. Balkema.
- Lanham, L. W. (1978). South African English. In L. W. Lanham, & K. P. Prinsloo (Eds.), *Language and Communication Studies in South Africa*. (pp. 138-165). Cape Town: Oxford University Press.
- Lanham, L. W. (1982). English in South Africa. In R. Bailey, & M. Gorlach (Eds.), *English as a World Language* (pp. 324-352). Cambridge: Cambridge University Press.
- Lanham, L. W. (1996). A History of English in South Africa. In V. De Klerk (Ed.), *Focus on South Africa* (pp. 19-34). Amsterdam: John Benjamins.
- Lanham, L. W., & Macdonald, C. (1979). *The Standard in South African English and its Social History*. Heidelberg: Julius Groos Verlag.
- Lanham, L. W., & Prinsloo, K. P. (Eds.). (1978). *Language and Communication Studies in South Africa*. Cape Town: Oxford University Press.
- Lanham, L. W., & Traill, T. (1962). *South African English Pronunciation*. Johannesburg: Witwatersrand University Press.
- Lass, R. (1988). Vowel Shifts, great and otherwise: Remarks on Stockwell and Minkova. In D. Kastovsky, G. Bauer, & J. Fisiak (Eds.), *Luick Revisited: papers read at the Luick Symposium at Schloß Liechtenstein 15, 1985* (pp. 395-410). Tübingen: Gunter Narr Verlag.
- Lass, R. (1990). A 'standard' South African vowel system. In *Studies in the pronunciation of English: a commemorative volume in honour of A.C. Gimson*. (pp. 272-285). London/New York: Routledge.
- Lass, R. (1995). South African English. In *Language and Social History: studies in South African Sociolinguistics* (pp. 89-106). Cape Town: David Philip.
- Lass, R. (2002). South African English. In R. Mesthrie (Ed.), *Language in South Africa*. (pp. 104-126). Cambridge: Cambridge University Press.

- Lass, R. (2004). South African English. In R. Hickey (Ed.), *Legacies of Colonial English: Studies in Transported Dialects* (pp. 363-386). Cambridge: Cambridge University Press.
- Lass, R., & Wright, S. (1985). The South African chain shift: Order out of chaos? In R. Eaton, O. Fischer, W. Koopman, & F. Van der Leek (Eds.), *Papers from the 4th International Conference on English Historical Linguistics* (pp. 137-161). Amsterdam: John Benjamins.
- Lass, R., & Wright, S. (1986). Endogeny vs. Contact: 'Afrikaans Influence' on South African English. *English World Wide*, 7 (2).
- Lehmann, W. P. (1962). *Historical Linguistics: an Introduction*. New York: Holt, Rinehart and Winston.
- Lennig, M. (1978). Acoustic measurement of linguistic change: The modern Paris vowel system. PhD Thesis: University of Pennsylvania .
- Lobanov, B. (1971). Classification of Russian Vowels Spoken by Different Speakers. *Journal of the Acoustical Society of America*, 49, 606-608.
- Luick, K. (1921). *Historische Grammatik der englischen Sprache*. Leipzig: C. H. Tauchnitz.
- Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change*, 19, 1-25.
- Malherbe, V. C. (1990). *Krotoa called 'Eva' - a Woman Between*. Cape Town: UCT African Studies.
- Martinet, A. (1952). Function, structure and sound change. *Word* 8, 1-32. In P. Baldi, & R. Werth (Eds.), *Readings in Historical Phonology (1978)* (pp. 121-59). University Park: The Pennsylvania State University Press.
- Martinet, A. (1955). *Economie des changements phonétiques*. Berne: Francke.
- McCormick, K. (1989). English and Afrikaans in District Six: a sociolinguistic study. PhD Thesis: University of Cape Town.
- McCormick, K. (1995). Code-Switching, code-mixing and convergence in Cape Town. In R. Mesthrie, *Language and Social History: Studies in South African Sociolinguistics*. (pp. 193-208). Cape Town: David Philip.
- Mesthrie, R. (1992). *English in Language Shift: The History, Structure and Sociolinguistics of South African Indian English*. Cambridge: Cambridge University Press.
- Mesthrie, R. (1993). English in South Africa. *English Today*, 9 (1), 27-37.

- Mesthrie, R. (1995). South African Indian English: from L2 to L1. In R. Mesthrie (Ed.), *Language and Social History: Studies in South African Sociolinguistics*. (pp. 251-64). Cape Town: David Philip.
- Mesthrie, R. (1997). A Sociolinguistic Study of Topicalisation Phenomena in South African Black English. In M. Görlach, & E. W. Schneider (Eds.) *Englishes Around the World: Carribean, Africa, Asia, Australasia*. Amsterdam/Philadelphia: John Benjamins.
- Mesthrie, R. (2002). From second language to first language: Indian South African English. In R. Mesthrie (Ed.), *Language in South Africa* (pp. 339-355). Cambridge: Cambridge University Press.
- Mesthrie, R. (2010). Socio-phonetics and Social Change: deracialisation of the GOOSE vowel in South African English. *Journal of Sociolinguistics*, 14 (1), 3-33.
- Mesthrie, R. (2012a). Second-Language Varieties: English in Africa - a diachronic typology. In A. Bergs, & L. J. Brinton (Eds.), *English Historical Linguistics: An International Handbook* (Vol. 2). Amsterdam: De Gruyter Mouton.
- Mesthrie, R. (2012b). Ethnicity, substrate and place: the dynamics of Coloured and Indian English in five South African cities in relation to the variable (t). *Language Variation and Change*, 24 (3), 371-396.
- Mesthrie, R. (2014). A lesser globalisation: A sociolexical study of Indian Englishes in Diaspora, with a primary focus on South Africa. In M. Hundt, & D. Sharma (Eds.), *English in the Indian Diaspora* (pp. 171-186). Amsterdam/Philadelphia: John Benjamins.
- Mesthrie, R., & Chevalier, A. (2014). Sociophonetics and the Indian Diaspora: the NURSE vowel and other selected features of South African Indian English. In M. Hundt, & D. Sharma (Eds.), *English in the Indian Diaspora* (pp. 85-104). Amsterdam/Philadelphia: John Benjamins.
- Mesthrie, R., Chevalier, A., & Dunne, T. (2015). A Regional and Social Dialectology of the BATH vowel in South African English. *Language Variation and Change*, 27, 1-30.
- Mesthrie, R., Chevalier, A., & McLachlan, K. (2015). A perception test for the deracialisation of middle class South African English. *Southern African Linguistics and Applied Language Studies*, 33(4), 391-409.
- Millar, B., Vonwiller, J., Harrington, J., & Dermody, B. (1994). The Australian National Database of Spoken Language. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing 2*. (pp. 67-100). Perth: The University of Western Australia.

- Milroy, J. (1980). Lexical alternation and the history of English: evidence from an urban vernacular. In E. Traugott (Ed.), *Papers from the 4th International Conference on Historical Linguistics*. Amsterdam and Philadelphia: John Benjamins.
- Milroy, J. (1992). *Linguistic Variation and Change*. Oxford: Blackwell.
- Milroy, J., & Harris, R. (1980). When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World Wide*, 1, 199-210.
- Milroy, L., & Gordon, M. J. (2003). *Sociolinguistics: method and interpretation*. Oxford: Blackwell.
- Mitchell, A. G., & Delbridge, A. (1965). *The Pronunciation of English in Australia*. Sydney: Angus and Robertson.
- Morreira, K. (2012). Black South African English: a sociophonetic study. PhD Thesis: University of Cape Town.
- Mufwene, S. (2008). Colonisation, population contracts, and the emergence of new language varieties: A response to Peter Trudgill. *Language in Society*, 37, 254-258.
- O'Grady, C., & Bekker, I. (2011). Dentalisation as a Regional Indicator in General South African English: an Acoustic Analysis. *Southern African Linguistics and Applied Language Studies*, 28 (1), 77-88.
- R Core Team. (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.Rproject.org/>.
- Rampton, B. (1995). *Crossing: Language and Ethnicity among Adolescents*. London: Longman.
- Roach, P. (1983). *English Phonetics and Phonology*. Cambridge: Cambridge University Press.
- Robinson, A. (1994). *The British English Example Pronunciation Dictionary*. Cambridge University.
- Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite. <http://fave.ling.upenn.edu>.
- Sankoff, D. (1980). Sociolinguistic and syntactic variation. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey. Vol IV: Language: The Socio-cultural Context*. Cambridge: Cambridge University Press.
- Sapir, E. (1921). *Language*. New York: Harcourt Brace.

- Schlemmer, L. (1976). English-speaking South Africans today: identity and integration into the broader national community. In A. de Villiers (Ed.), *English-speaking South Africa Today* (pp. 91-136). Cape Town: Oxford University Press.
- Schmied, J. (1991). *English in Africa: an Introduction*. London: Longman.
- Schneider, E. W. (2003). The dynamics of New Englishes. From identity construction to dialect birth. *Language*, 79, 233-81.
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Schneider, E. W. (2011). *English around the World: an introduction*. Cambridge: Cambridge University Press.
- Schrijver, P. (2011). The High German Consonant Shift and Language Contact. *Studies in Slavic and General Linguistics*, 38, 217-249.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23, 193-229.
- Spencer, B. (2011). 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.
- Spencer, J. (1971). West Africa and the English language. In J. Spencer (Ed.), *The English Language in West Africa* (pp. 1-34). London: Longman.
- Statistics South Africa*. (n.d.). Retrieved 2014 from SuperWEB: <http://interactive.statssa.gov.za/superweb/login.do>
- Stockwell, R. P., & Minkova, D. (1988). The English Vowel Shift: problems of coherence and explanation. In G. Kastovsky, G. Bauer, & J. Fisiak (Eds.), *Luick Revisited: papers read at the Luick Symposium at Schloß Liechtenstein 15, 1985* (pp. 355-394). Tübingen: Gunter Narr Verlag.
- Strelluf, C. (2014). "We have such a normal, non-accented voice": a sociophonetic study of English in Kansas City. PhD Thesis: University of Missouri-Columbia.
- Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional Variable Importance for Random Forests. *BMC Bioinformatics*, 9 (307), URL <http://www.biomedcentral.com/1471-2105/9/307>.
- Strobl, C., Boulesteix, A., Zeileis, A., & Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, 8 (25), URL <http://www.biomedcentral.com/1471-2105/8/25>.
- Sweet, H. (1888). *A History of English Sounds*. Oxford: Clarendon Press.

- Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden/Oxford: Wiley-Blackwell.
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Language Variation and Change*, 24, 135-178.
- Thomas, E. R. (2003). Secrets revealed by Southern vowel shifting. *American Speech*, 77, 150-170.
- Thomas, E. R. (2011). *Sociophonetics: An Introduction*. United Kingdom: Palgrave Macmillan.
- Toefy, T. (2014). Sociophonetics and Class Differentiation: a study of working- and middle-class English in Cape Town's Coloured Community. PhD Thesis: University of Cape Town.
- Toefy, T. (forthcoming). The Status of the KIT-split in Coloured South African English.
- Torgersen, E., & Kerswill, P. (2004). Internal and external motivation in phonetic change: Dialect levelling outcomes for an English vowel shift. *Journal of Sociolinguistics*, 8 (1), 23-53.
- Torgersen, E., Kerswill, P., & Fox, S. (2006). Ethnicity as a source of changes in the London vowel system. In F. Hiskens (Ed.), *Language Variation - European Perspectives. Selected papers from the Third International Conference on Language Variation in Europe (ICLaVE 3), Amsterdam, June 2005*. (pp. 249-264). Amsterdam: John Benjamins.
- Trager, G. L. (1940). One phonemic entity becomes two: The case of 'short a'. *American Speech*, 15, 255-258.
- Trudgill, P. (1999). Norwich: endogenous and exogenous linguistic change. In P. Foulkes, & G. Docherty, *Urban Voices: Accent Studies in the British Isles* (pp. 124-140). London: Arnold.
- Trudgill, P. (2004). *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
- Trudgill, P., Gordon, E., & Lewis, G. (1998). New-dialect formation and Southern Hemisphere English: The New Zealand short front vowels. *Journal of Sociolinguistics*, 2 (1), 35-51.
- Van den Heever, C. M., & Wissing, D. (2000). Tswana first language interference on English vowels. In D. Wissing (Ed.), *Proceedings of the Workshop on Black South African English* (pp. 34-39). Cape Town: Linguistics Society of South Africa.

- Van Der Pas, B., Wissing, D., & Zonneveld, W. (2000). Parameter resetting in metrical phonology: the case of Setswana and English. *South African Journal of Linguistics, Supplement 28*, 55-87.
- Van Rooy, B. (1995). Word-final devoicing by Tswana and Afrikaans speakers of English: Second language interference or a universal tendency? *Paper presented at the Linguistics Society of South Africa Conference, 5-7 July*. University of Port Elizabeth/Rhodes University.
- Van Rooy, B. (2000). The consonants of BSAE: current knowledge and future prospects. *South African Journal of Linguistics, Supplement 38*, 55-87.
- Van Rooy, B. (2004). Black South African English: phonology. In E. W. Schneider, K. Burrige, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (Vol. 1: Phonology, pp. 943-952). Berlin: Mouton de Gruyter.
- Van Rooy, B. (2010). Social and linguistic perspectives on variability in World Englishes. *World Englishes*, 29 (1), 3-20.
- Van Rooy, B. (2014). Convergence and endonormativity at Phase 4 of the Dynamic Model. In S. Buschfeld, T. Hoffmann, M. Huber, & A. Kautzsch (Eds.), *The Evolution of Englishes: the Dynamic Model and beyond*. (pp. 21-38). Amsterdam/Philadelphia: John Benjamins.
- Van Rooy, B., & Terblance, L. (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, B., & Van Huyssteen, G. B. (2000). The vowels of Black South African English. *South African Journal of Linguistics, Supplement 38*, 15-33.
- Warren, P. (2006). Oops, I've done a futt: quality and quantity in a New Zealand vowel contrast. *Te Reo*, 49, 125-143.
- Watson, C., Maclagan, M., & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change*, 12, 51-68.
- Webb, V. (1983). On a South African English vowel system. *Stellenbosch Papers in Linguistics*, 10, 134-164.
- Wells, J. (1982). *Accents of English Volumes 1-3*. Cambridge: Cambridge University Press.
- Welsh, F. (1998). *A History of South Africa*. London: Harper Collins.
- Wiebesiek, L. (2007). Addressing the 'Standard English' Debate in South Africa: the case of South African Indian English. MA Thesis: University of KwaZulu Natal.

- Wileman, B. (2011). Regional variation in South African English: a socio-phonetic comparison of young white speakers in Cape Town and Durban. MA Thesis: University of Cape Town.
- Wileman, B. (in progress). A sociophonetic investigation of differences in voice quality as a indicator of ethnicity for young, middle-class South African English speakers. [working title].
- Wilmot, K. (2014). "Coconuts" and the middle class: Identity change and the emergence of a new prestigious English variety in South Africa. *English World Wide*, 35 (3), 306-337.
- Wolfram, W., & Schilling-Estes, N. (1998). *American English: Dialects and Variation*. Malden, Massachusetts: Blackwell.
- Wood, T. (1987). Perceptions of and attitudes towards varieties of English in the Cape Peninsula, with particular reference to the 'Coloured Community'. MA Thesis: Rhodes University.
- Woods, N. J. (1997). The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics*, 1, 95-125.
- Woods, N. J. (2000). New Zealand English across the generations: An analysis of selected vowel and consonant variables. In A. Bell, & K. Kuiper (Eds.), *New Zealand English* (pp. 84-110). Wellington: Victoria University Press.
- Yuan, J., & Lieberman, M. (2008). Vowel Acoustic Space in Continuous Speech: An Example of Using Audio Books for Research. *Proceedings of the conference on "The First International Workshop on Cataloguing and Encoding of Spoken Language Data" at CatCod 2008*.