An Acoustic Analysis of New Zealand English Vowels in Auckland

BY

Brooke Chantel Ross

Abstract

This study presents an acoustic analysis looking at phonetic diversity in Auckland. New Zealand English is often characterized by a lack of regional variation; however, this claim has been made without considering Auckland, New Zealand's largest city. Over the last 30 years there has been increased migration to New Zealand, specifically to Auckland. In 35% of Auckland's suburbs, no ethnic group represents more than 50% of the population. In addition, many speakers were born overseas, and many more have grown up using different varieties of English as the spoken norm. In this study, 40 New Zealand English speakers from three suburbs in Auckland (Mt. Roskill n= 14, Papatoetoe, n=13, Titirangi, n=13) were recorded. For our young group (n=33) the participants were aged between 18 and 25 years, and each suburb was evenly split between male and female participants. Speakers were either New Zealand born or arrived in the country under the age of seven. Our older group (n=7) were female speakers, all New Zealand born, and aged between 40 and 70 yrs. Vowels which had sentence stress were identified and extracted, and formant values were calculated at the vowel target. All formant tracks were hand checked. Over 8000 monophthong tokens and 4000 diphthongs were analysed in this study. Whilst no differences were found between young speakers from different suburbs, there were age effects. Further, speech from the young Auckland speakers was noticeably different to findings from other studies on New Zealand English. Most notably monophthongs TRAP and DRESS were lower than expected. In addition the first targets of the diphthongs FACE and GOAT have risen, and PRICE has fronted, for younger speakers from all suburbs. The thesis concludes discussing the implications of the results.

Table of Contents

Chapter 3: Background - Auckland Voices Project   24

Chapter 4: Background – Method   31

Chapter 5: Method   43

Chapter 6: Results   57

Chapter 7: Discussion of Results   91

List of Figures

List of Tables

Acknowledgements

First, I need to thank everyone involved in the Auckland Voices Project for giving me the opportunity to undertake this research, without you all this would not have been possible. In particular, thank you to Helen Charters and my supervisor Miriam Meyerhoff for encouraging my involvement in this project and believing it was something I could undertake.

The biggest thanks of all must go out to my supervisor Catherine Watson for spending countless hours every week teaching me about acoustic analysis, vowels, New Zealand English, and anything in between. You have been an indispensible fountain of knowledge over the past year, and I cannot thank you enough. It has been a pleasure being taught by you.

I would also like to extend a big thank you to my supervisor Elaine Ballard for constantly being on hand to help me structure my ideas and thoughts. I could not have completed this thesis without your wonderful advice and encouragement.

Thank you to everyone from our weekly speech group meetings for listening to my progress every week and keeping me on track. Thank you for your advice, and for just generally being a kind group of people. In particular, I must thank Hywel Stoakes for dedicating hours to troubleshooting various coding problems. My work would have been significantly more arduous without your help.

Huge thanks also to my friend Melissa Irvine for constantly listening to my complaints, stupid questions, and for just generally providing sound linguistic advice over the past year. Thank.

Special thanks to my partner in crime Jacob Pynenburg for your support and patience. You have kept me grounded and stress-free throughout, and it has made this process infinitely more enjoyable.

Finally I have to thank my amazing family and friends who have provided me with the support necessary to write this thesis. I quite literally could not have done this without your help.

Chapter I
Introduction

In New Zealand you can take seven hours to drive from Wellington to Auckland, and not be able to hear any difference in the English that is spoken when you arrive. (Bauer, 1994)

It is commonly claimed that there is very little or no regional variation in the New Zealand accent, as evidenced by the quote above. The only exception to this is in the rhotic dialect found in the Southland region of the South Island of New Zealand (Bartlett, 1992). As far as vowels are concerned though, studies undertaken in cities such as Wellington and Christchurch have found no variation. Strangely, however, Auckland, New Zealand's largest city has never been the focus of a linguistic analysis. Some studies have looked at small sub-groups of Auckland speakers for evidence of specific varieties of New Zealand English such as Pasifika English (Gibson & Bell, 2010; Starks, Gibson & Bell, 2015), but none have considered the vowel spaces of the wider Auckland community. This is a major problem for two reasons: First, Auckland is home to over a third of New Zealand's English speaking population. Second, over the past 30 years immigration to Auckland has resulted in massive demographic change. Given the distinctive profile of English in Auckland compared with the rest of New Zealand, it seems pertinent to investigate variation in the vowel spaces of Aucklanders. This thesis, therefore, presents an acoustic vowel analysis of 40 New Zealand English speakers in Auckland in order to answer the question: Is English in Auckland the same or different to regional New Zealand English?

In Chapter II necessary background information about New Zealand English is presented, as well as a discussion about the processes of language change in diverse cities. Chapter III presents background information about the database of speakers used for this project. Chapter IV

provides a discussion about the thorough data preparation processes undertaken in this acoustic study. Chapter V presents the acoustic analysis method undertaken. Chapter VI presents the results of the acoustic analysis. Finally, Chapter VII discusses the results of the acoustic analysis and concludes.

Chapter II
Background: New Zealand English

This section aims to provide necessary background information on New Zealand English. First a brief history and typology of New Zealand English is provided; followed by a small discussion regarding the significance of New Zealand English. Then, given this thesis presents an acoustic vowel analysis, a comprehensive discussion of New Zealand English monophthongs and diphthongs is provided. This is followed by a small discussion of two varieties of New Zealand English which are relevant to this study: Māori English and Pasifika English. Finally sociolinguistic factors of importance to this study are identified and discussed. Important factors relate both to the linguistic diversity of Auckland and how change might function within a diverse metropolis, and known sociolinguistic processes present in New Zealand.

Throughout this thesis the vowels of New Zealand English are represented using lexical sets from Wells (1982). Phonemic IPA labels largely based on Wells (1982) are also occasionally used for monophthongs, with the exception of KIT where /ɪ/ is used instead of /ə/. Diphthongs are referred to by their lexical set equivalent throughout. Both are given in Figure 1 below. Phonemic as opposed to phonetic labels are used because this thesis is largely interested in variation. Using phonetically accurate labels for labelling would obscure vowel changes. In addition, labelling vowel centroids for each group by their nearest IPA approximate would quickly become confusing.

	Monophthongs		**Diphthongs**
FLEECE	iː		FACE
DRESS	e		PRICE
KIT	ɪ		CHOICE
TRAP	æ		MOUTH
STRUT	ʌ		GOAT
START	aː		NEAR
LOT	ɒ		SQUARE
FOOT	ʊ		
THOUGHT	ɔː		
GOOSE	uː		
NURSE	ɜː		

Figure 1. Lexical sets and phonemic IPA symbols based on Wells (1982b).

2.1 New Zealand English Background

Typologically, New Zealand English (NZE) is a variant of English from the South East of England brought to New Zealand by migrants during large scale 19th century migration (Wells, 1982b). From the 1840's immigrants mainly from England, but also from Ireland, Scotland and Australia made their way to New Zealand. There is minimal demographic information regarding specific cities of origin for the earliest New Zealand immigrants. It is known, however, that many were from London, in addition to other cities such as Cornwall, Devon, and Oxfordshire (Gordon, 1998). This means that New Zealand English phonology is essentially a variant of south-east English systems (Bauer, 1994). It uses the same consonant inventory, and, with only a couple exceptions, the same vowel inventory. As a result of similar settlement histories, New Zealand English also shares many features with other Southern Hemisphere varieties of English, particularly Australian English (Trudgill, Gordon, Lewis and Maclagan, 2000), for instance, in being non-rhotic.

Despite similarities with other Southern Hemisphere varieties there are several vowel features which are often said to characterize NZE. Most notably, New Zealand English has a very retracted and centralized KIT vowel, close/high realizations of TRAP and DRESS, and a merged variant of the diphthongs NEAR and SQUARE. In addition the NURSE vowel is notably rounded. These vowel features are discussed in more detail in section 2.2. Another noteworthy feature of NZE is the use of very dark or vocalized /l/ in syllable coda position (Wells, 1982b; Maclagan, 1982; Horvarth & Horvarth, 2001).

2.1.1 Class Variation

Outside of regional contrasts the only variation amongst New Zealand dialects is one based somewhat on class. New Zealand English can be classified into two categories, Broad or Cultivated (Gordon & Deverson, 1998). This is based on a similar classification in Australian English. Cultivated has traditionally been used to refer to speech nearest to (although not identical to) RP, and broad is used to refer to speech that differs from RP - often associated with the speaking style of rural New Zealand farmers (Gordon & Deverson, 1998, pp. 136). It is important to note, rather than existing as two separate varieties, speakers tend to fall somewhere on a continuum, with broad on one end of the spectrum and cultivated on the other (Gordon & Deverson, 1998). More recent work has also considered the possibility of distinct Māori and Pasifika varieties of New Zealand English; these are discussed further in section 2.3.

2.1.2 Significance of New Zealand English

The study of New Zealand English has played an important role in theories of dialect formation and contact. It is an intriguing case study because of its short linguistic history and relative geographic isolation. European immigration to New Zealand spans only 180 years meaning there is fairly comprehensive chronological information regarding its development. Both written and recorded historical records give insight into how processes of sound change have resulted in the modern New Zealand English dialect. Written records identifying the emergence of a new variety of English date back to a newspaper article from 1889 commenting on New

5

Zealand English pronunciation by Sam McBurney (Turner, 1967). Written records continue throughout the early 20th century, largely through complaints regarding the degradation of the English language in New Zealand from prestigious varieties spoken in Britain (Gordon, 1998, pp. 65). Despite the condescending nature of these records, early commentary does identify the early existence of many features which would later come to define New Zealand English. These are, for instance, the retraction and lowering of KIT (Wall, 1948; Gordon, 1998), and the opening and fronting of STRUT (Turner, 1967); or the broad New Zealand English pronunciation of PRICE with a rounded first target (Gordon, 1998). In addition to written records, recorded speech dating back as far as 2nd generation European immigrants also provides valuable data on the development of New Zealand English. The Origins of New Zealand English (ONZE) project investigates recordings collected in the 1940s of conversations with New Zealanders born in the second half of the 19th century (Maclagan & Gordon, 2004). This group largely represents the first generation of "native New Zealand English" speakers (pp. 42). Investigation of these recordings has contributed greatly to theories of dialect formation and dialect contact (Trudgill, Gordon, Lewis & Maclagan, 2000). Because the development of early New Zealand English happened largely in isolation, ONZE speakers provide a wealth of information on how the processes of dialect mixture and leveling function. This is further discussed in section 2.4.3.

2.2 New Zealand English Vowels

Despite using the same vowel inventory, realizations of New Zealand English vowels are vastly different from their English counterparts. The earliest acoustic analysis of the New Zealand vowel space was completed in 1982; it suggested the following changes from RP (Maclagan, 1982). First, New Zealand English has a triangular vowel space with three point vowels, much like Australian English, rather than a vowel quadrilateral with four point vowels. Notable vowel differences from RP include raised TRAP, DRESS, THOUGHT and LOT vowels; as well as centralization of GOOSE, and raising of NURSE, to the extent that the two largely overlap in the F1/F2 vowel space. Maclagan also noted lowered and centralized FOOT and KIT vowels; and the diphthongization of FLEECE, GOOSE and DRESS. Finally comments were made about the similarity

of NEAR and SQUARE. Also noted were the coarticulation effects of dark /l/, resulting in retraction for all vowels. This study is important as it marks the first quantifiable evidence of how NZE vowels differ from other varieties of English.

New Zealand English has undergone some changes since the earliest acoustic studies. The following sections will discuss in detail relevant features of the modern New Zealand English vowel space. Many vowels in NZE are involved in ongoing changes which are particularly relevant to this study, so where necessary these changes will be described as well. Figure 2 below shows the vowel space of modern New Zealand English monophthongs.
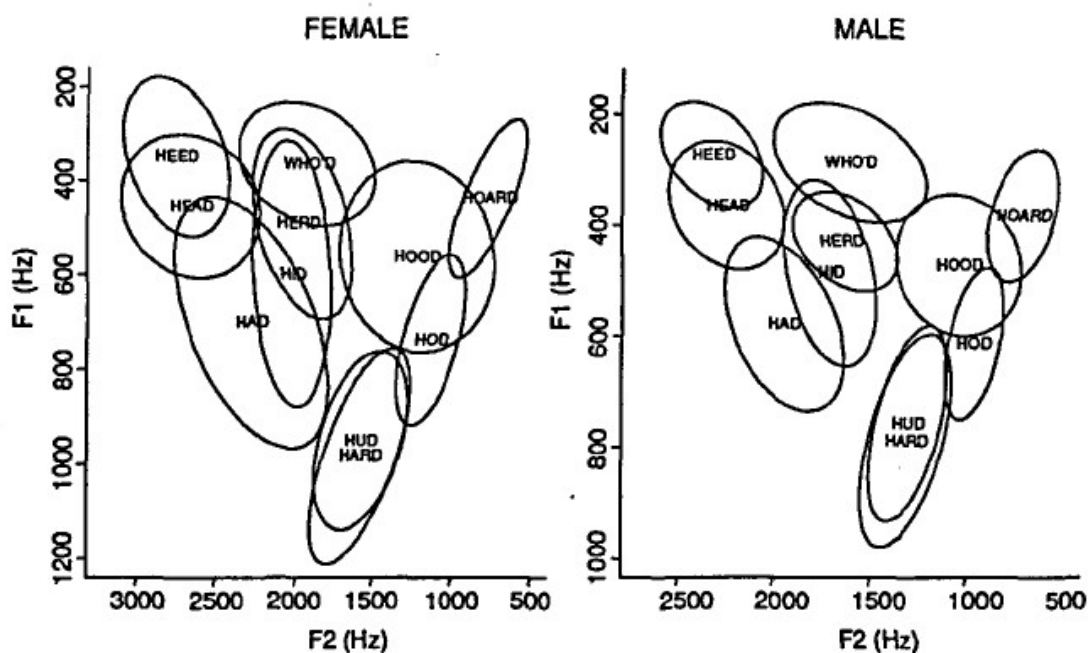
Figure 2. Ellipse plots of NZE monophthongs in F1/F2 space for female *(left)* and male *(right)* speakers. Reprinted from *An acoustic comparison between New Zealand and Australian English vowels,* by Watson, Harrington & Evans, 1998.

Perhaps the most linguistically important change currently underway in New Zealand English is the ongoing chain shift involving high front vowels; this is comprehensively discussed in Section 2.2.1.; NURSE is covered in section 2.2.2.; Section 2.2.3 covers the back vowels THOUGHT, LOT and FOOT; and START and STRUT are covered in Section 2.2.4.; Finally, section 2.2.5 discusses New Zealand English diphthongs, including the merger of NEAR and SQUARE.

2.2.1 Front Vowels

The most interesting vowels in the New Zealand English vowel system are the front vowels, FLEECE, KIT, DRESS and TRAP. Since the late 1970's it has been claimed that these vowels are involved in an ongoing chain shift (Bauer, 1979; 1992; Trudgill, Gordon & Lewis, 1998). First identified by Bauer (1987, 1992) the assertion was made that New Zealand English short vowels had innovatively moved compared to traditional "English". TRAP and DRESS had raised, and KIT had retracted. In contrast, in the Australian short front vowel system TRAP and DRESS also rose, but instead of retracting, KIT became even closer. At first there was some debate about whether these front vowels were involved in a push chain or drag chain movement. Proponents of the push-chain theory assert that the chain shift began with the TRAP vowel, moving from [æ] towards [ɛ]. In order to maintain a distinction between DRESS and TRAP, DRESS was pushed upwards towards [e]. This then resulted in DRESS encroaching on the vowel space of KIT, which instead of raising towards FLEECE retracted in the direction of schwa. This is the now the most commonly accepted theory on how this chain shift functions.

The most recent studies on New Zealand English front vowel raising have suggested a continuation of this chain shift where DRESS has raised even further towards FLEECE (Maclagan and Hay, 2007; Warren, 2017). In fact there is some suggestion that for younger speakers the vowel spaces of DRESS and FLEECE overlap entirely, or even that in some cases DRESS is high and more fronted than FLEECE. Initially it was thought that these speakers were using a length contrast to distinguish these vowels (Warren, 2017). More recently however it has been shown that there are negligible length differences between the two vowels, and this cannot be used as a reliable measure to distinguish between them. Instead both Maclagan & Hay (2007) and

Warren (2017), propose that as DRESS rises to overlap with FLEECE, FLEECE is diphthongizing and for young speakers has a considerable onglide. For many speakers this diphthongization is increasingly becoming the method by which these phonemes are distinguished.

2.2.1.1 Australian English Front Vowel Shift

Front vowel raising is not exclusive to New Zealand English. It is also found in Australian English. The chain shift resulting in the Australian English vowel system functioned slightly differently to New Zealand English, however. Instead of DRESS raising resulting in KIT retraction, in Australian English the raising of DRESS resulted in further raising of KIT towards FLEECE (Cox & Palethorpe, 2001). In Australian English KIT and FLEECE now form a long short pair. This is commonly considered the most salient difference between Australian and New Zealand English. Recent studies suggest, however, that a reversal of this front vowel shift maybe occurring in southern varieties of Australian English (SAusE) (Cox & Palethorpe, 2007; Cox & Palethorpe, 2008; Elvin, Williams & Escudero, 2016). Cox and Palethorpe (2008) studied the vowel spaces of 30 speakers of Australian English from Sydney. They found lowered TRAP, and evidence for lowering of DRESS amongst these speakers. They also note that DRESS lowering is accompanied by the lowering of the NURSE vowel amongst these speakers. A recent study of 19 Australian English speakers from Western Sydney also supported these claims, finding DRESS and TRAP vowels comparatively lower than previous studies (Elvin et al., 2016). This is unusual because previous work suggests sound change reversals don't occur (Bybee, 2003). An earlier acoustic study, Cox & Palethorpe (2001), of 27 speakers from Sydney found a lowered TRAP vowel, but no lowering of DRESS. This suggests that the reversal of front vowel movement in SAusE began with TRAP, and is thus a drag chain. Cox and Palethorpe (2008) note that both the push-chain that initiated vowel raising and the reversal of this change discussed here were initiated by the movement of TRAP. So far acoustic studies of the front vowel space in NZE have only discovered vowel raising. Given the close relationship between southern hemisphere varieties, it is important to note that a reversal of front vowel shift in NZE is possible.

2.2.2 NURSE Vowel

Hay et al. (2008) identify the NURSE vowel as one of the most characteristic sounds of New Zealand English. Historically NURSE in New Zealand English is realized as a close-mid vowel (Watson, Harrington & Evans, 1998; Hay, Maclagan & Gordon, 2008). A kinematic analysis by Watson et al. (1998) also showed that NURSE is rounded for both New Zealand and Australian English speakers. Recent studies have shown, however, that the NURSE vowel appears to have undergone a sound change. Maclagan, Watson, Harlow, King, & Keegan (2017) investigate NURSE further. Analyzing a group of bilingual Māori and NZE speakers, they found that the NURSE vowel had become closer. These speakers produced NURSE vowels with formants values near those of GOOSE. Surprisingly, despite their similar formant values, lengths and lip-rounding, the two vowels did not appear to be merging. Finally, there is a possibility that raised NURSE is linked to DRESS raising in New Zealand English. Studies on SAusE have suggested that as DRESS lowers NURSE is lowered (Cox & Palethorpe, 2008), implying a relationship between the two vowels in Southern Hemisphere varieties of English.

2.2.3 Back Vowels

The back vowels of New Zealand English are THOUGHT, LOT and FOOT. All three vowels are rounded, although some speakers produce unrounded variants of LOT and FOOT (Allan & Starks, 2000). THOUGHT, like most other dialects of southern hemisphere English, is noticeably raised (Allan & Starks, 2000). In comparison with RP the NZE FOOT is traditionally lowered (Allan & Starks, 2000), although recent research suggests it is raising (Warren, 2017; Watson, Palethorpe & Harrington, 2004). The LOT vowel is somewhat unremarkable, but like other dialects of southern hemisphere English is also fairly raised compared to RP (Allan & Starks, 2000).

Recent work suggests that the FOOT and THOUGHT may be undergoing changes in New Zealand English. For instance, in a study tracking the vowel movements of three speakers over 40~ years Watson, Palethorpe & Harrington (2004) showed that THOUGHT has risen and backed, and the FOOT vowel had also risen. There has also been some suggestion that FOOT and GOOSE might now form a long/short pair (Bauer & Warren, 2004, as cited in Warren 2017). Most recently,

however, it has been suggested that for younger speakers FOOT is fronting towards GOOSE (Warren 2017). Either way the movement of the FOOT vowel does not seem to impact the formant values of THOUGHT or GOOSE vowels (Warren, 2017). In addition, much of the debate around the cardinal location of FOOT is likely related to its short duration and infrequent use in speech. Its short duration means FOOT formant values are heavily influenced by surrounding consonants, making it hard to accurately identify.

2.2.4 START *and* STRUT

Another notable feature of the New Zealand vowel system is the fronting of the STRUT and START vowels in comparison to Received Pronunciation (RP). First mentioned by McBurney in his 1887 commentary on New Zealand English, it seems as though at some point in the history of NZE formation STRUT lowered, and START fronted. The two now form a tense/lax or short/long pair in New Zealand English (Watson el al., 1998; Warren, 2017). In fact a recent study suggests that these vowels form the only legitimate short/long contrast in New Zealand English (Warren 2017).

2.2.5 Diphthongs

New Zealand English has seven diphthongs as shown in Figure 3[1].
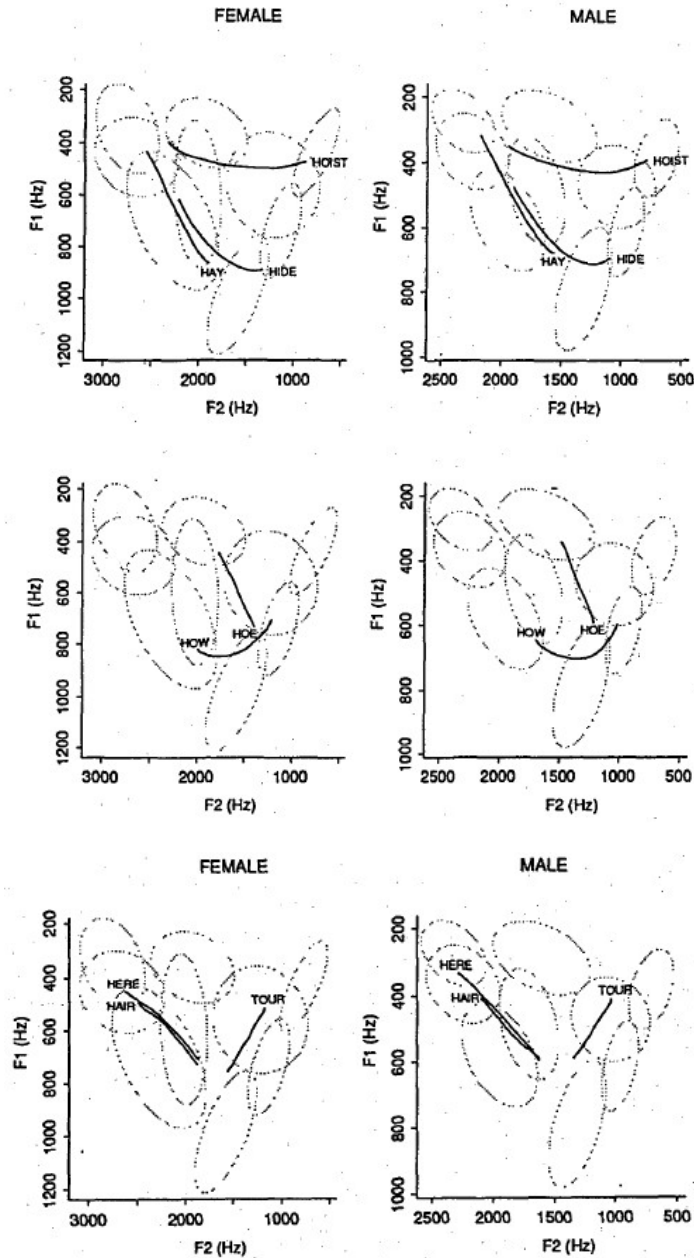
Figure 3. Trajectories of NZE Diphthongs in F1/F2 space for female *(left)* and male *(right)* speakers. Reprinted from *An acoustic comparison between New Zealand and Australian English vowels,* by Watson, Harrington & Evans, 1998.

---

[1] This Figure also shows the diphthong CURE which we do not consider in this study.

The merger of falling diphthongs NEAR and SQUARE is the most well documented diphthong change in NZE (Maclagan, 1982; Holmes & Bell, 1992; Maclagan & Gordon, 1996; Watson et al., 1998; Watson, Maclagan & Harrington, 2000). It is well documented that sometime in the past 100 years the first target of SQUARE [eə] rose and the first target of NEAR [iə] fell slightly, so that the two are now phonetically identical for most New Zealand English speakers (Watson et al., 2000). It is thought that the movement of the short front monophthongs in NZE was the catalyst for the merger of these diphthongs, particularly DRESS raising which corresponds to the first target of SQUARE (Watson et al., 2000; Maclagan & Gordon, 1996). Essentially, as realizations of DRESS became increasingly close, so did the first target of SQUARE, resulting in confusion with NEAR. Initially this movement towards NEAR was somewhat stigmatized and some speakers produced hyper-corrected variants of SQUARE, for both NEAR and SQUARE, resulting in some initial confusion about which direction the merger was moving (Maclagan and Gordon, 1996). Now the direction of the merger is clear, and it is essentially complete, with most young NZE speakers unable to tell the difference between the two diphthongs (Watson et al., 2000). While some young speakers still produce variants equivalent to both NEAR and SQUARE, they are no long phonemically distinct (Watson et al., 2000).

Other Diphthongs

While the merger of NEAR and SQUARE is the most notable diphthong feature of NZE, it is by no means the only unique feature of the diphthong space. Other NZE diphthongs also differ from RP, although they are largely similar to their Australian counterparts. One difference between NZE and AE diphthongs, however, is the suggestion that the first targets of the front-rising diphthongs FACE and GOAT in NZE have raised with their monophthongal counterparts (Watson, et al. 1998). For instance, Allan & Starks (2000) note that FACE in southern hemisphere dialects of English has a lowered first target[2] compared to RP. Watson et al. (1998), however, found FACE was more raised than the corresponding Australian English vowel. This raising of FACE could correspond to the raised TRAP in New Zealand English, which is close to the first target of FACE. Similarly CHOICE has a raised first target in southern hemisphere varieties of English (Allan &

---

[2] For an in depth discussion about vowel targets see Section 4.1.1

Starks, 2000). The first target of CHOICE in NZE is even higher than the corresponding AE target. According to Watson et al. (1998), this corresponds to raised THOUGHT in NZE. Although it should be noted, the raised first targets of the diphthongs are less extreme than the corresponding raised monophthongs (Watson et al., 1998).

One of the most iconic diphthongs in NZE is MOUTH. According to Allan & Starks (2000), MOUTH in southern hemisphere varieties of English has a raised first target. This first target has also retracted in RP to create an even greater contrast. There is also a non-standard variant of MOUTH produced by some speakers of broad NZE which has a further raised and notably fronted first target (Britain, 2008). At the extreme the first target of this variant of MOUTH can approach DRESS giving IPA realizations such as [ɛɪ] or [ɛ⁹] as opposed to [ạ̈ö] (Britain, 2008; Allan & Starks, 2000). Britain (2008) notes that this variant is difficult to analyze in acoustic studies because it is highly stigmatized and speakers tend not to produce it in interview settings.

Finally the PRICE and GOAT vowels in NZE also differ from RP. The first target of PRICE is lowered and retracted (Allan & Starks, 2000; Watson, et al. 1998). Allan & Starks (1998) also note that some speakers produce a variant with a rounded first target, but again this is stigmatized and hard to capture in an interview environment. As for GOAT, Allan & Starks (2000) note that southern hemisphere varieties have lowered first target compared to RP. Furthermore they note that the second target in NZE can be rounded or unrounded. They give two IPA variants [ɐö] and [ɐ̞ɤ]. Watson et al. (1998) note that NZE realizations of GOAT are more raised than AE variants.

2.3 Pasifika and Māori English

Although most scholars agree that there is little dialectal variation between regions, there is evidence for emerging dialects within some New Zealand communities. Māori English a variation of NZE influenced by the Māori language is well documented (Bell, 2000; Szakay, 2007; Maclagan, King, & Gillon, 2008; Warren & Bauer, 2008). In addition, recent research has

searched for evidence of a NZE variety influenced by Pacific languages dubbed "Pasifika English" (Gibson & Bell, 2010; Starks, Gibson & Bell, 2015). Given the linguistic diversity of Auckland, it is highly likely that speakers of these varieties of NZE exist within Auckland communities, and hence relevant features could emerge in our data. Here a brief summary of proposed differences between NZE and Pasifika/Māori English is given.

2.3.1 Māori English Vowels

Māori English as a variety of New Zealand English has been commented on in literature since the middle of the 20th century (Maclagan, King, & Gillon, 2008). Although initially condemned as an uneducated dialect in the 1960's and 1970's, it has since been recognized by academics as an important marker of social identity spoken by many Māori and Pākehā in New Zealand (Warren & Bauer, 2008). It is likely that Māori English developed from the transference of features from Māori to English amongst the earliest Māori learners of English during the 19th century (Bell, 2000). These features were then passed down through generations of Māori even as proficiency in the Māori language declined. Today a large percentage of New Zealanders who identify as Māori learn English as their first and often only language; which means Māori English can serve as an important marker of identity (Warren & Bauer, 2008).

Unfortunately little of the work on Māori English focuses on vowels. Work instead focuses on supra-segmental features such as syllable timing (Szakay, 2007). This is because phonological features of Māori English have been notoriously hard to identify in acoustic studies. Warren & Bauer (2008), comment on the difficulty of capturing Māori English in a linguistic setting:

> Because of the very nature of Māori English, getting good recordings of this variety in formal settings, in a Pākehā institution (a university) and with Pākehā researchers is difficult. None of the recordings provided here is completely prototypical, even when we have Māori people speaking to each other without Pākehā people present. (pp. 83)

15

Nevertheless, some features of Māori English vowels have been identified. Bell (1997), mentions in passing that acoustic evidence suggests Māori English has fronted and lowered DRESS compared to Pākehā NZE. Bell (2000) and Bell (1997), also provide acoustic evidence that Māori English speakers have a less centralized KIT than Pākehā English speakers. Presumably due to the corresponding short /i/ vowel in Māori. Speakers also often use KIT rather than schwa in unstressed syllables (Bell, 1997). As for the long vowels, Warren & Bauer (2008) based on acoustic evidence that FLEECE is "probably" less diphthongized than Pākehā NZE. In addition the THOUGHT is notably lower. Bell (2000) also mentions GOOSE fronting as a possible feature of Māori English. There is also the possibility that NURSE is more fronted and rounder than Pākehā English (Warren & Bauer, 2008). Diphthongs do not tend to differ from Pākehā New Zealand English (Warren & Bauer, 2008).

2.3.2 Pasifika English Vowels

Despite migrating from many different nations and communities with vastly different languages and cultures, socio-economic factors often mean that in New Zealand, many Pacific communities live in close quarters. In parts of Auckland these groups to some extent form a larger overall community based on their shared Pacific heritage (Starks, Gibson & Bell, 2015). This interaction has led to scholars suggesting the existence of a shared "Pasifika English" variety of NZE emerging from these communities (Gibson & Bell, 2010; Starks, Gibson & Bell, 2015). Although the features of this dialect can vary depending on the cultural background of the speakers, some shared or more common features have been identified.

Recent research by Starks, Gibson & Bell (2015), considers impressionistic evidence from the South Auckland suburb, Manukau. Based on data collected to explore language maintenance amongst speakers of Cook Island, Samoan, Tongan, Niuean and Māori descent, they discuss the influence of Pacific Island communities on the varieties of English spoken there. Another study analyzes features in the animated TV comedy 'Bro Town' which caricaturizes Pacific and Māori stereotypes, and therefore speaking styles (Gibson and Bell, 2010). Although the linguistic features performed in this show are exaggerated and largely stereotyped, they supposedly give

some indication as to features that might be expected in these Pacific influenced varieties of New Zealand English. These studies suggest the possibility of the following features arising from contact with Pasifika languages: First, a somewhat raised KIT vowel; second, reduced diphthongization of FLEECE and GOOSE; in addition lowered DRESS, and for some speakers a lowered and retracted TRAP. There is also some evidence suggesting STRUT is raised compared to START. Finally there is evidence suggesting a greater than normal rate of monophthongization of NZE diphthongs amongst these speakers.

Furthermore research by Thompson, Starks, & Watson (2008) suggests some vowel features from Niuean speakers of New Zealand English, from an acoustic analysis of around 1000 vowel tokens. The Niuean community is the fourth largest Pacific community in New Zealand (Thompson, Starks, & Watson, 2008), and hence this data might provide some insight into features of a "Pasifika English" variety. This study found that for Niuean New Zealand English speakers TRAP, DRESS and KIT were raised, in line with the Pasifika English studies above. In addition the Niuean speakers considered also had raised THOUGHT and LOT vowels. As for diphthongs the study found raised onsets for PRICE and MOUTH, as well as overall shortened monophthong compared to New Zealand English.

In addition to vowel features, studies also suggest other linguistic features unique to "Pasifika English". The phenomena of TH-fronting/DH-stopping where [θ] and [ð] become [f] and [v], or [t] and [d]; and intrusive /r/ are the most significant, however (Gibson & Bell, 2010). Here vowels are of most interest, so these will not be covered in detail.

2.4 Sociolinguistic Background

The following sections discuss sociolinguistic processes which are relevant to New Zealand English and Auckland. Section 2.4.1 discusses language change in metropolises followed by a discussion in Section 2.4.2 about whether Auckland constitutes a metropolitan city. Following this, Section 2.4.3 talks about dialect formation in early New Zealand English and how it relates to this study. Section 2.3.4 talks about the relationship between gender and sound change in New Zealand English, and finally section 2.3.5 proposes some research questions based on this sociolinguistic background.

2.4.1 Languages in Metropolises

Many large cities across the world are becoming increasingly multicultural. In recent years studies have begun to explore how cultural diversity in large metropolises impacts language. In cities such as London, new dialects are emerging as a result of language contact between "British English" and the multitude of languages spoken by recent immigrants (Cheshire et al., 2011, 2013). This work on Multicultural London English (MLE) has inspired an onslaught of related studies considering variation in large cities across the world such as Sydney (Sydney Speaks Project, n.d.) and Paris (Sneddon, 2015). While Auckland may not quite be of the same scale as London, it is extremely diverse and substantially larger than any other city in New Zealand.

It is important then to consider how language change might function in a large city such as Auckland. Cheshire et al. (2013) looked at Central London and a variety of English they dubbed "Multicultural London English" or MLE. The study focused on two suburbs of London, one central London multi-cultural suburb, and one largely homogenous outer London suburb. They found that rather than any particular features of substrate languages being of significance, it was the language contact setting itself that drove the development of the dialect (pp. 14). They noted huge inter-speaker linguistic variation, and a great amount of flexibility in the linguistic norms of the speakers in the more diverse suburb. As a result of the many different inputs speakers were able to be more creative and innovative with their speech (pp. 10). Applied to

Auckland this might mean in more diverse areas more variation from New Zealand English norms is expected.

2.4.2 Auckland as a Metropolis

The 2013 census places Auckland's population at around 1.4 million people, or a third of NZ's population (Statistics NZ, 2013a). This is almost triple the size of New Zealand's next largest region, Wellington. In addition Auckland is significantly more diverse than any other region in the country. At the 2013 census, 39.1 % of people in the Auckland region reported being born overseas compared with 25.2% across all of New Zealand (Statistics NZ, 2013b). There are also less monolingual English speakers in Auckland compared with the rest of New Zealand. 30.6% of Aucklanders claim to speak more than one language compared with 20.2% for the rest of New Zealand (Statistics NZ, 2013b).

Despite the diversity and increasing multiculturalism of Auckland, there is little research involving the speech of Auckland participants. Most studies looking at the vowel spaces of New Zealand English speakers have taken place using participants from regional New Zealand. These studies have generally reached a consensus that there is little regional variation amongst New Zealand English speakers. Given this, it will be intriguing to investigate how the historical homogeneity of New Zealand English interacts with Auckland's super-diverse rich linguistic environment.

2.4.3 Dialect Formation in New Zealand

This section considers dialect contact and dialect formation processes in a New Zealand context. If cultural and linguistic diversity turn out to be driving factors behind language change in Auckland as Cheshire et al. (2011, 2013) predicts, then having an understanding of how dialect formation and dialect contact function will become important. Trudgill, Gordon, Lewis and Maclagan (2000), discuss dialect mixture and new dialect formation in the early stages of New Zealand English. This is of interest both because of the local context and because they suggest a framework for the emergence of new dialects from an original dialect mixture. They

propose three stages of dialect mixing based on the ONZE data briefly discussed in Section 2.1.2, these are:

1. Rudimentary leveling.
2. Extreme variability.
3. Focusing and the emergence of a New Dialect.

The first stage *rudimentary leveling* refers to the initial contact adult speakers of different dialects make in a location. Speakers must accommodate to each other in order to be understood, and therefore many obscure features of individual dialects are leveled out.

The second stage *extreme variability* refers to the quantity of inter-speaker variation in the accents of the children of recent immigrants. Trudgill et al. (2000) note that children usually acquire the dialects of their peers, but in dialect mixture situations there is not a specific dialect to acquire. As a result adults play a greater role in the development of these children's dialects. The adults speak a wide range of different dialects themselves, and therefore the children have a wide variety of linguistic models to pick and choose features from (pp. 305). This results in unique and unusual combinations of linguistic features, and greater inter-speaker differences. Following this variability there is further dialect levelling similar to the first stage, which Trudgill et al. (2000) consider a sub-stage of this second stage.

The final stage consists of *focusing and the emergence of a new dialect*. This is where a new dialect can first be identified. Speakers lose their variability and focus in on specific features of a new dialect. The reasons as to why speakers focus on particular features and not others is a complicated issue. The effects, however, are a new homogenous dialect. This final stage is clearly demonstrated in the lack of regional variation in New Zealand English.

Trudgill et al. argue for this leveling process to explain how features of minority dialects can survive dialect mixture situations. For instance if dialect formation simply consisted of majority forms swamping other features as suggested by Lass (1990) then New Zealand English should sound a lot more like the South of England varieties used by the majority of early immigrants.

Although New Zealand English is phonologically a variant of South East English dialects, there are many features of these dialects which did not survive the initial dialect mixture.

It is interesting to consider how these processes might apply to the Auckland data considered in the following chapter. It seems that in less diverse areas, a dialect should remain stable, undergoing predictable natural language change without pressure from other languages and dialects. In an extremely diverse suburb such as Mount Roskill, however, where many participants are children of immigrants who are not native New Zealand English speakers, more parallels are drawn with the ONZE data. Here the second stage process of extreme variability seems particularly relevant. Combined with theory from Cheshire et al. (2011, 2015) regarding language variation in metropolises, large amounts of inter and intra-speaker variation might be expected in our two more diverse suburbs discussed in Section 3.2.

There are key differences between the Auckland data considered in this study and the ONZE data, however. First, ONZE speakers as the children of immigrants in the mid-late 19th century would have had very little external linguistic influence outside of their immediate community. Auckland participants, however, are recruited from suburbs within a large city that are not entirely isolated from one another. While all were raised and most schooled in the relevant suburb, many young speakers now in their early 20s travel almost daily outside of their suburb either to work or university, and are likely to interact with speakers from across Auckland in these environments. In addition, Modern New Zealand is not isolated from the rest of the world as the ONZE speakers were. Speakers with diverse backgrounds are likely to have frequent contact with friends and family across the world through online mediums, and the increasing ease of world travel. This is particularly true in Auckland, New Zealand's international hub. While the city itself is immensely diverse, this diversity is further increased by the recent boom of international tourism. Almost every international traveler arriving in New Zealand, by obligation has to spend some time in Auckland. This likely exposes Aucklanders to even more diversity than that which is expressed by its population statistics. Any dialect formation processes underway in diverse Auckland suburbs, such as Mount Roskill, are not going to simply be the result of dialect mixture in a linguistically diverse isolated community – constant external

factors are also likely to play a role. This begs the question of how external influences might impact the processes of leveling, variability and focusing which Trudgill et al. (2000) identify.

2.4.4 Gender and Sound Change in New Zealand

Sociolinguistic work on New Zealand English has suggested males and females differ in how they adopt sound changes. In New Zealand, multiple studies have shown that young women are sound change innovators, leading non-stigmatized sound changes, but tend to be more conservative with stigmatized changes. They also tend to lead sound changes which go unnoticed by the general populace. This is in line with the principles of sound change proposed by Labov (1994). For instance Maclagan, Gordon & Lewis (1999), found women use stigmatized forms of diphthongs PRICE & MOUTH less than male counterparts (stigmatized change). In addition, women seem to be leading front vowel raising processes in New Zealand English (Maclagan et al., 1999; Maclagan & Hay, 2007); particularly the raising of DRESS and diphthongization of FLEECE (unnoticed change). Finally, Gordon & Maclagan (2001) found that women were also more advanced in the merger of NEAR and SQUARE (non-stigmatized change). This incites the distinct possibility that any sound changes which emerge in this data are likely to have progressed further for female speakers, unless these sound changes are stigmatized.

2.5 Research Questions

The primary goal of this study is to investigate whether the New Zealand English vowel space is undergoing change in Auckland. Considering sociolinguistic factors such as diversity, language contact and gender provides some narrower questions which can guide analysis:

1. Is English in Auckland the same or different to regional New Zealand English?

2. Is there a relationship between gender and language change in Auckland? Previous studies suggest that young women tend to lead sound change in New Zealand, is this true in Auckland?

3. Are more diverse areas of Auckland more linguistically variable as suggested by Cheshire et al. (2013), or does variation and dialect contact function differently in Auckland?

4. Are there age effects between young and older speakers in Auckland?

5. Given the previous questions, are young women from diverse suburbs the most variable speakers of New Zealand English?

Chapter III

Background: Auckland Voices

3.1 Auckland Voices Project

In order to investigate the questions proposed above, a database of Auckland speakers is of course required. Fortunately such a database exists. This acoustic study uses data collected for the "Auckland Voices" project. Auckland Voices is a Marsden funded project investigating the impact that increased migration and diversity is having on English in Auckland (Meyerhoff, 2017). As part of this project, interviews approximately one hour in length were collected with participants from three Auckland suburbs. These interviews were collected by trained research assistants or project supervisors, with a sociolinguistic framework in mind. This meant that interviews were conducted in interviewees' homes or other places they felt comfortable. The aim of the interviewer was to encourage natural conversation with the participant, while avoiding the formality that interview situations tend to entail. Each participant was also asked to read a set passage aloud towards the end of the interview, this passage is given in Appendix A.

3.1.1 Read Speech

Two types of data are analyzed in this thesis. The first is from a read passage speakers were asked to read aloud during their interviews. This data is the main focus of the acoustic study described in Chapter 5. Although read speech is more conservative than conversational speech, using a read passage controls some variability. This means that any differences identified between groups and speakers are likely to be more reliable. The read speech performed is taken from a study Holmes and Boyce (1991), "Variation and change in New Zealand English: a social dialect study". It was designed to be socio-culturally appropriate for a New Zealand English speaking audience. It also contains certain linguistic variables they were interested in studying.

These variables are:

1. Every NZE vowel possible before a voiceless obstruent
2. Every NZE vowel possible before a syllable final /l/
3. possible contexts for linking /r/ and intrusive /r/
4. Some specific words which might have variant pronunciations.
   Holmes and Boyce (1991, pp.28)

This study is not particularly concerned with any of these variables. They are important to note, however, because an abnormally high number of occurrences of a variable could impact our results. For instance, syllable final /l/ is known to have a major impact on vowel formant values. This is discussed in more detail in Section 4.1.5.

3.1.2 Conversational speech

Read speech is likely to be more conservative than a casual vernacular in any situation (Labov, 1972). While interviews provide speech that is more natural than a read passage, in any formal interview situation participants are still likely to perform more conservatively than their casual vernacular (Labov, 1972; Meyerhoff, Schleef & Mackenzie, 2015). For instance, we have already mentioned the difficulty in capturing natural speech data from Maori New Zealand English speakers (Warren & Bauer, 2008). The interviews analyzed in this study were collected with a sociolinguistic framework in mind, however. This entails that much thought was put into how to reduce the formality of the interview setting. Interviews were conducted at a comfortable, quiet, location of the participants choosing. This tended to be at their home, but also could be at a nearby library, café, or university, for instance. This ensured a less formal setting where participants felt comfortable using their casual vernacular. Interviews were also conducted by members of the relevant community where possible, to further facilitate comfortable conversation. Interviewers were given a list of topics and questions to memorize in order to facilitate conversation. For example interviewers would ask participants to describe their identity or discuss changes they have seen in Auckland. They were aware, however, that the

main purpose of the interview was to facilitate casual conversation, and not simply to elicit answers to specific questions. This ensured the conversations recorded were as casual and comfortable as could be in an interview setting.

3.2 Participant Selection

This section gives necessary background information about participant selection for the Auckland Voices Project. This participant selection shapes the acoustic analysis performed in this thesis as seen in the revised research questions in Section 3.3 and the data selection in Section 5.1.

First, participants had to meet the following requirements to be considered for the study: All participants had to be speakers of New Zealand English. Data was collected from equal numbers of younger and older speakers. Younger speakers could be no older than 25, and must have moved to New Zealand before age seven. Older speakers had to be over 40. Interviews were collected with participants from 3 suburbs across Auckland - Titirangi, Mount Roskill, and Papatoetoe. These suburbs were chosen as they represent different vastly different Auckland communities. The location of each suburb is shown in Figure 4 below. Within each suburb interviews were conducted with equal numbers of male and female speakers. Interviews were then transcribed using the annotation software ELAN (Sloetjes & Wittenburg, 2008) by trained research assistants. The following sub-sections discuss the sociolinguistic relevance of each suburb for the Auckland Voices Project.

3.2.1 Titirangi

The first suburb, Titirangi, represents a largely homogenous community in the West Auckland area known as the Waitākere Ranges. Its residents are largely New Zealand European, and tend to be from wealthier families. The Waitākere Ranges Local board reports that 78.8 of the usual resident population identify as European compared to 59.3 percent across Auckland (Auckland Council, 2014a). This number is even higher in the smaller area considered for this study. For instance at the 2013 census, 92.7 percent of the usual resident population of the "area unit"

South Titirangi, as defined by NZ Stats, identified as European (Statistics NZ, n.d.). According to Statistics NZ (2013b), compared to the rest of Auckland, more Titirangi residents are also born in New Zealand (28% born overseas, compared with 39.1% across Auckland). Cheshire et al. (2013) suggests that language change in metropolises is often driven by increased cultural diversity. Given the previous claims about the homogeneity of New Zealand English, and the homogeneity of speakers in the Titirangi region, participants from Titirangi are expected to produce a variety of New Zealand English which deviates least from previous studies on New Zealand English.
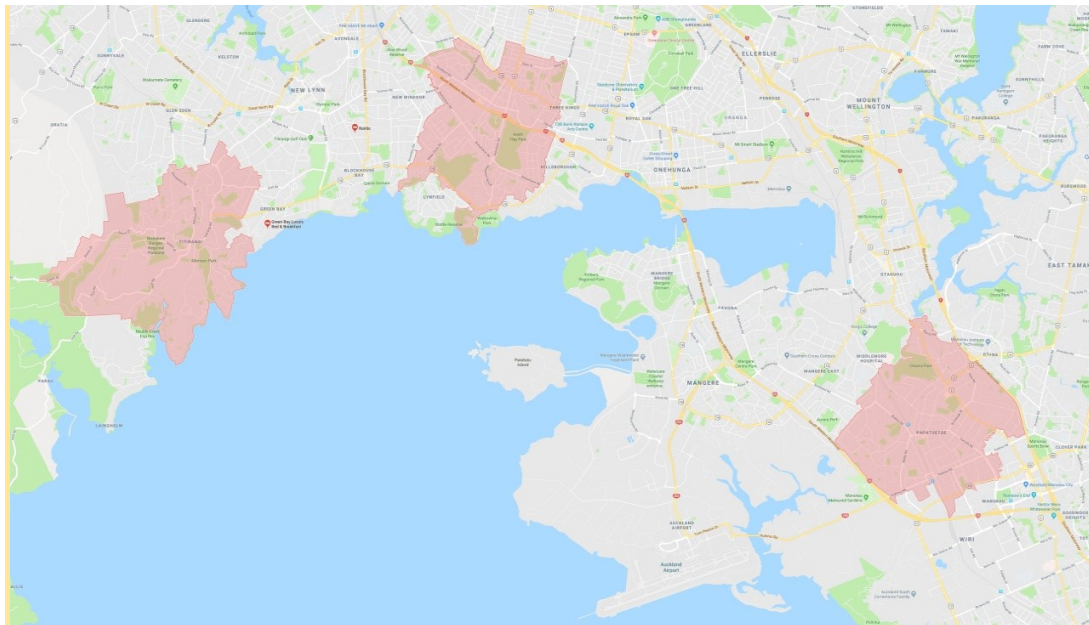
Figure 4. Map showing the location of the three suburbs considered by the Auckland Voices project. Suburbs highlighted in red: from left to right, Titirangi, Mount Roskill, and Papatoetoe (Google Maps, 2018).

3.2.2 Papatoetoe

The second suburb considered for this study, Papatoetoe, represents a more diverse community in South Auckland. According to the Ōtara-Papatoetoe Local Board only 20.7 percent of the usual resident population identify as European, compared with 59.3 across the rest of Auckland. 45.7 percent of the population identify as Pacific Peoples which is significantly higher than any other suburb studied here, and in Auckland overall (14.6%); while 15.6 percent of the usual resident population identify as Māori. A further 30.9 percent of Papatoetoe residents identify as "Asian". Although Papatoetoe has the lowest European population in our dataset, Mount Roskill is home to a larger number of individuals "born overseas". According to the Auckland Council, 42% of Ōtara-Papatoetoe residents are born overseas (Auckland Council, 2014b), compared with 49% of Puketāpapa residents (Auckland Council, 2014c). Where the speakers in our Mount Roskill data set are largely the children of recent migrants, speakers from Papatoetoe often come from well established Pacific communities who have resided in New Zealand for two or more generations. In addition Pacific countries arguably have more contact with New Zealand English and similar southern hemisphere dialects. For these reasons, while we expect speakers from Papatoetoe to be more diverse than speakers from Titirangi, we expect less variation than found in the Mount Roskill group.

Samoan is the second most common language in Ōtara-Papatoetoe, being spoken by 17.2% of the population. Given recent interest in varieties of "Pasifika English" (Gibson & Bell, 2010; Starks, Gibson & Bell, 2015), if there is evidence of a "Pasifika" variety of NZE we might expect it to be more prevalent in speakers from Papatoetoe. Certainly many recent studies have focused their research on identifying Pasifika English in this or nearby suburbs such as Manukau (Gibson & Bell, 2010). Papatoetoe also has a larger number of Te reo Māori, Hindi, and Tongan speakers than Auckland as a whole (Auckland Council, 2014b).

3.2.3 Mount Roskill

The final suburb analyzed, Mount Roskill, was selected as it has undergone rapid diversification in the past 20 years. Cheshire et al. (2013) suggests, as diversity increases language variation increases. It is expected therefore, that the most linguistically diverse speakers will be young Mount Roskill participants.

As of the 2013 census, the Puketāpapa local board reports that 38.0 percent of usual residents identify as European. While 44.2% identify as Asian, and 15.9% identify as Pacific Peoples (Auckland Council, 2014c). The Puketāpapa local board however, consists of not only Mount Roskill, the suburb considered for this study, but also the nearby suburbs of Mount Eden and Mount Albert. These suburbs while approximately the same size as Mount Roskill, are notably less diverse. For instance the "area unit" *Akarana* - as defined by Statistics NZ, corresponds to approximately the centre of Mount Roskill. Statistics NZ reports that in this area 19.6 percent of residents identify as European, 28.6 percent identify as Pacific Peoples and 48.8 percent identify as Asian (Statistics NZ, n.d.). This is significantly different to the stats for the entire Puketāpapa local board.

The largest ethnic group that Mount Roskill residents identify with is "Asian". NZ Statistics combines a wide range of different cultures and communities into this label, however. According to the Puketāpapa local board the largest groups represented are people of Indian and Chinese descent, although this is also not particularly specific. There are also many other cultures present in the Mount Roskill population encompassed by the label "Asian",including but not limited to Sri Lankan, Filipino, Korean, Malaysian and Japanese (Auckland Council, 2017). This is without considering the diversity present within the two larger ethnic groups labeled by NZ Stats as simply Indian or Chinese. Aside from English Mount Roskill is home to large numbers of Hindi, Northern Chinese, Samoan, and Tongan speakers.

3.3 Revised Research Questions

Considering the Auckland Voices Data allows the research questions posed in Section 2.5 to be refined based on how sociolinguistic factors from Section 2.4 relate to the Auckland Voices data. The first and primary question remains the same:

1.   Is English in Auckland the same or different to regional New Zealand English?

2.   Are there gender effects between male and female speakers? Given previous research, do young women lead sound change in Auckland?

3.   Are there significant cross suburb differences in vowel spaces supporting Cheshire et al. (2013), or does variation and dialect contact function differently in Auckland?

   3a. Furthermore, given we have two diverse suburbs to consider for analysis, is there a difference between how recent (Mount Roskill) and established (Papatoetoe) migrant communities respond to dialect contact processes.

4. Are there age effects between speakers under 25 and speakers over 40?

5. Given the previous questions, are young Mount Roskill women the most variable speakers of New Zealand English in Auckland?

Chapter IV

Background: Methodology

4.1 Data Preparation Processes

Some of the data preparation process undertaken in this thesis might seem arduous and time consuming given there are plenty of automatic services available for use. If an automatic segmentation system exists, for instance, why hand check vowel and word boundaries? This section discusses in detail the necessity of cautious and careful data preparation and advocates for a careful combination of automatic and manual data preparation.

4.1.1 Stressed Vowels

Unstressed vowel tokens are unreliable for acoustic analysis for several reasons. Therefore, while preparing the data, it was decided that only vowels tokens in prosodically stressed positions would be considered for vowel analysis. First of all, unstressed vowels are heavily centralized towards schwa, while prosodically stressed vowels fall towards the outer edges of the vowel quadrilateral (Harrington & Cassidy, 1999). This centralization is also referred to as "vowel undershoot", referring to vowels which fail to reach their targets (Harrington, 2010). Although vowel undershoot can also occur with stressed vowels, it occurs significantly more often with unstressed tokens. This centralization obscures differences between vowels, and makes sound change harder to identify. In addition unstressed vowel tokens are often heavily coarticulated, especially when surrounded by approximants such as /r/, /l/ or /w/ (Harrington, 1999). Vowel undershoot and coarticulation are often conflated, especially in unstressed tokens. Coarticulation specifically refers to the simultaneous production of two sounds, however, resulting in one influencing the other. For example in Figure 5 below the low second formant of the [w] lowers the high 2nd formant of the first target of SQUARE. This obscures the first target of SQUARE which would normally be marked at the peak of F2, towards the start of the vowel. Unstressed vowel tokens are particularly susceptible to coarticulation effects because they tend to have reduced duration.
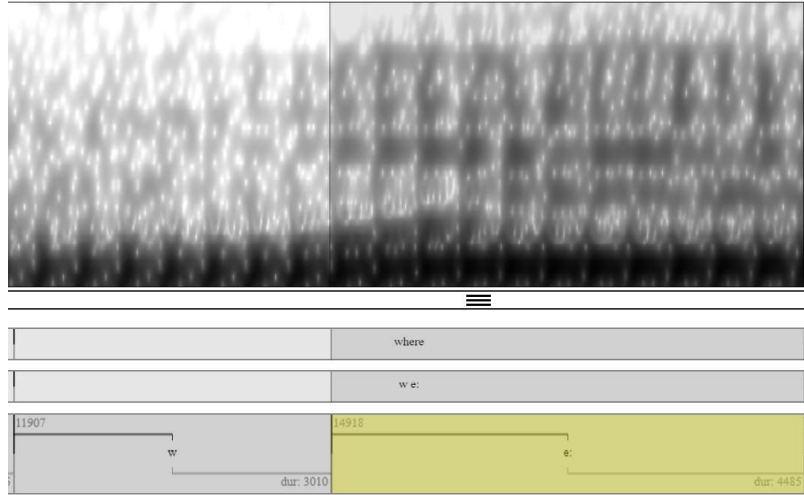
Figure 5. Spectrogram of vowel undershoot caused by coarticulation, in this instance SQUARE preceded by the approximant /w/.

While coarticulation can impact the identification of vowel targets, it also makes accurately labeling vowel boundaries challenging. In many extreme cases, vowels preceding or following voiceless sounds often disappear, for example Figure 6 below.

In Figure 6 below, a spectrogram of two unstressed words, the GOOSE vowel in "used" and the schwa in "to" contain barely any phonetic information. Rather than dark bands indicating first and second formants, the influences of the surrounding voiceless sounds results in little to no voicing of the vowel. These tokens provide no formant values and therefore no information for analysis.
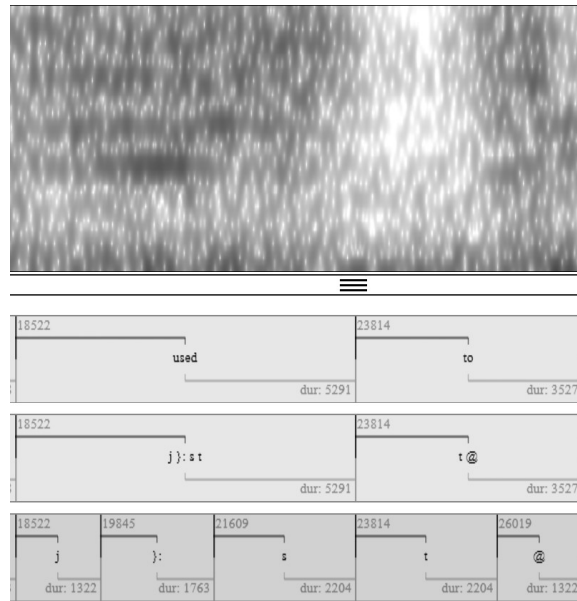
Figure 6. Spectrogram showing coarticulation effects of neighbouring voiceless consonants on unstressed vowel tokens.

Given the reasons above, unstressed vowels in many cases, don't provide any particularly useful phonetic information, and if anything obscure differences between sounds because of centralization. While some unstressed tokens might in fact be suitable for analysis, selecting tokens on a case by case basis is somewhat capricious. It is most appropriate to exclude tokens on a categorical basis. Therefore, the decision was made to exclude all unstressed tokens.

4.1.2 Correcting Phoneme and Word Boundaries

Although phonemes were automatically labeled using the WebMAUS automatic labeler software, English (NZ) service, (Kisler, Reichel, & Schiel, 2017), vowel tokens still required hand checking to ensure boundaries were accurate, and to correct any labeling errors. Automatic labeling is extremely useful and saves countless hours of work, but it is by no means perfect. Errors are common due to external factors such as recording quality and speaker style/variation, as well as internal software factors.

First, while WebMAUS saves immeasurable time by creating labels that would otherwise have to be hand-created; the phoneme and word boundaries it sets are often inaccurate. This is not to say that WebMAUS cannot identify sounds and parse them correctly. Rather, start times and end times provided by WebMAUS tend to be slightly askew from the boundaries visible in spectrograms. For instance James, Tian & Watson (2018), found that for clear speech recorded in a quiet environment (18dB ambient noise), 15% percent of word and phoneme boundaries required correcting. Given that this study considers vowel duration and vowel onglide, it is important that the vowel tokens we label have accurate boundaries. Hence, hand checking is necessary.

WebMAUS' segmentation capabilities become further compromised when certain external factors occur. For instance, the recordings used in this data set were recorded according to sociolinguistic guidelines to ensure the most natural speech possible. Interviews were largely held in participant's homes or other comfortable environments. Although these environments were as quiet as possible, they were by no means silent. There is always varying amounts of uncontrollable background noise in each recording, for instance birds chirping, refrigerator hums, unidentifiable static, children and so on. If the recording is too noisy the automatic segmentation can be compromised. This is also true of speaker volume. While interviewers checked microphone volume at the beginning of each recording, speaker volume often changes throughout speech. In a sociolinguistic interview setting it would be very distracting for an interviewer to be constantly adjusting microphone volume. Thus in some recordings there will be instances of clipping, or speakers will produce speech that is too quiet. Both of these things compromise webMAUS's segmentation abilities. Hand checking phoneme labels allowed the correction of errors caused by the casual interview settings required for this study.

In addition webMAUS struggles to accurately parse speech when certain linguistic phenomena occur; for instance, if a speaker pauses, or if a speaker lengthens a syllable for emphasis or simply while contemplating what to say next. Plus, not all of our participants speak with the exact standard New Zealand accent that the English (NZ) service of webMAUS is based on.

Segmentation errors therefore often occur due to this speaker variation; for instance, alternate pronunciations of words.

It should also be noted that the errors described above are even more abundant in our conversational speech. This is the result of numerous differences which occur between the two speaking styles. For instance, speakers pause more frequently. Spontaneous sentences often don't follow predictable syntactic structures, as sentences in the designed passage do. In addition abrupt speech rate changes often confuse the segmentation software. Finally, perhaps the biggest error trigger exclusive to read speech is contending with interviewers talking over speakers. Interviews were conducted in a conversational style so simultaneous speech by the participant and interviewer in common. WebMAUS cannot parse simultaneous speech, so labeling confusion occurs if the interviewer and participant happen to speak at the same time. This can be problematic, as if webMAUS mislabels a word or phoneme early in an utterance, it will often result in everything which follows being mislabeled. This can make large portions of an utterance unusable without hand correction.

Overall, while automatic segmentation software is extremely useful, it should be used with caution and not completely relied upon. Hand checking automatically segmented data produced by webMAUS ultimately improved the accuracy of the acoustic analysis performed here and hence made the results more reliable.

4.1.3 Formant Correction

In addition to hand checking labels, the formants of stressed vowels calculated in R (R Core Team, 2013), using `WRASSP` algorithm within `EmuR` on default settings (Winkelmann, Jaensch, Cassidy & Harrington, 2018), were hand checked and corrected where necessary. This again is essential as while automatic formant track calculation is useful - it is by no means perfect. Certain factors, largely recording quality and vocal tract differences between genders can impact the accuracy of formant calculation.

Common formant tracking errors include: The switching of first and second formant tracks for back and low vowels; third formants tracking along the second formant and second formants tracking along the first, while the first tracks at zero Hz or F0; Formants tracking along visible vowel harmonics, rather than formants; Formant tracker not locating a formant when one is clearly visible, and tracking along zero Hz. These errors can occur across the entirety of the vowel production, or only a portion. Errors tend to occur when certain features of the sound recording interfere with the formant trackers abilities, but also can simply be a byproduct of formant tracking software. For instance, in recordings such as AK-PY01, an interview with a young male participant that contains minimal background noise, formant correction was minimal. This is the ideal scenario. This is far from the norm in the Auckland Voices dataset, however.

There are two major factors which compromise formant accuracy. The first is recording quality. Many of the recording quality factors which impact automatic segmentation in Section 4.1.2 also cause problems for formant calculation. Our recordings were produced in the most casual sociolinguistic settings possible. As a result of this, in some interviews the acoustic signal can be somewhat compromised. Background noise can obscure formants and cause them to mistrack. For instance, in a study considering the F1 and F2 formant values of three Arabic vowels, Al-Tairi, Watson, & Brown (2016) found that, in noisy recordings, 38% of formant tracks had to be corrected. This might be as simple as F2 tracking as F3, or it may cause more complex issues. For example, in the Auckland voices dataset, recording AK-PY05 was accompanied by the constant sound of birds chirping in a nearby tree. The high pitch of the birds caused the formant tracker to constantly track formants approximately 200Hz above the formants visible on the spectrogram. If these formants had not been checked this would not have been noticed, and it might have been assumed that this speaker had significantly more raised vowels than actuality. Most frequently, however, noisy recordings simply result in rough rather than smooth formant tracks, where the formant track fluctuates in frequency along the formant, for instance Figure 7 below.
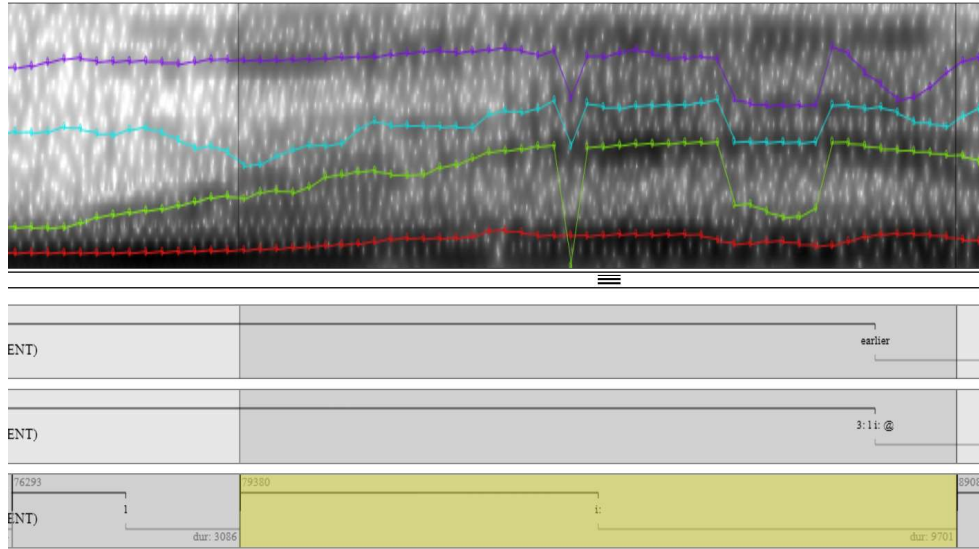
Figure 7. Spectrogram showing shaky formant tracks during production of a ꜰʟᴇᴇᴄᴇ vowel, as well as tracking errors caused by recording quality factors.

In Figure 7 while F1 and F2 are visible, at points they are not clearly defined. This results in formant tracks which to not transition smoothly from point A to point B. This is most problematic for diphthongs, where the formants are extracted between two targets along the formant track. Although where possible formants should not be corrected, leaving these uneven tracks results in diphthong plots which do not demonstrate the smooth transitions between two targets.

Speech volume also impacts the accuracy of formant tracks. While microphone volume is adjusted at the start of each interview, interviewers cannot predict amplitude fluctuations. In addition, having the interviewer constantly monitoring and adjusting the recording equipment during an interview would greatly detract from the casual setting these interviews aim at replicating. If a participant speaks too softly, often the formant tracker cannot locate F1 or F2. If F1 and F2 are visible on the spectrogram, or can be identified through looking at a spectral slice produced from the spectrogram, they can be corrected. This allows the vowel token to be used in analysis. Participants speaking at amplitudes too loud for the capability of the microphone

also cause problems for formant calculation. This phenomenon, known as clipping, occurs when the amplitude of the input signal is greater than the maximum value allowed by the recorder (Rosen & Howell, 1998). Often when clipping occurs, the vowel token has to be excluded as even visual examination cannot separate F1 and F2, for example Figure 8 below. If the first and second formant can be identified and are simply mistracking, these can fixed through hand correction. In Figure 8 below, F2 is clearly visible. F1 is buried under a mess of spectral information caused by clipping, however. Sometimes creating spectral slices of fixed time points in the vowel can help identify an F1 value, but this is not always the case. In this instance F1 could not be located and then token had to be excluded.
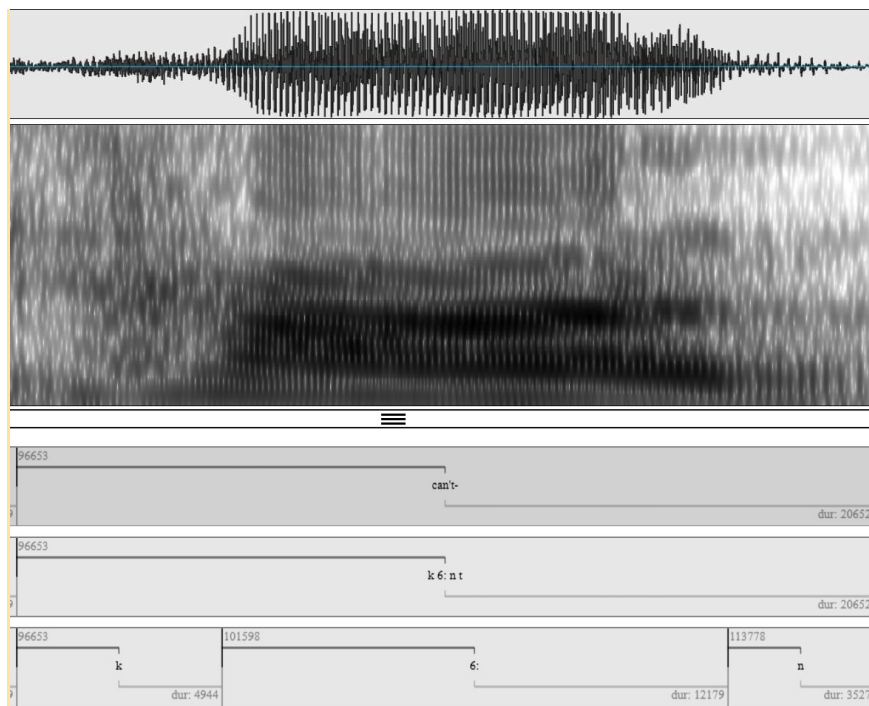
Figure 8. Spectrogram showing clipping in the word 'can't'. Evidence of clipping can be seen both spectrally, and in the wavform above the spectrogram.

Occasionally, when a speaker is not over clocking the microphone, but is still speaking rather loudly, individual vowel harmonics become visible on the spectrogram itself. In Figure 9 below

the formant tracker has mistracked a formant along one of these visible harmonics rather than the first or second formant. Usually the F1 and F2 will appear darker on the spectrogram than harmonics and these issues can be fixed. This is what has occurred in Figure 9 below.
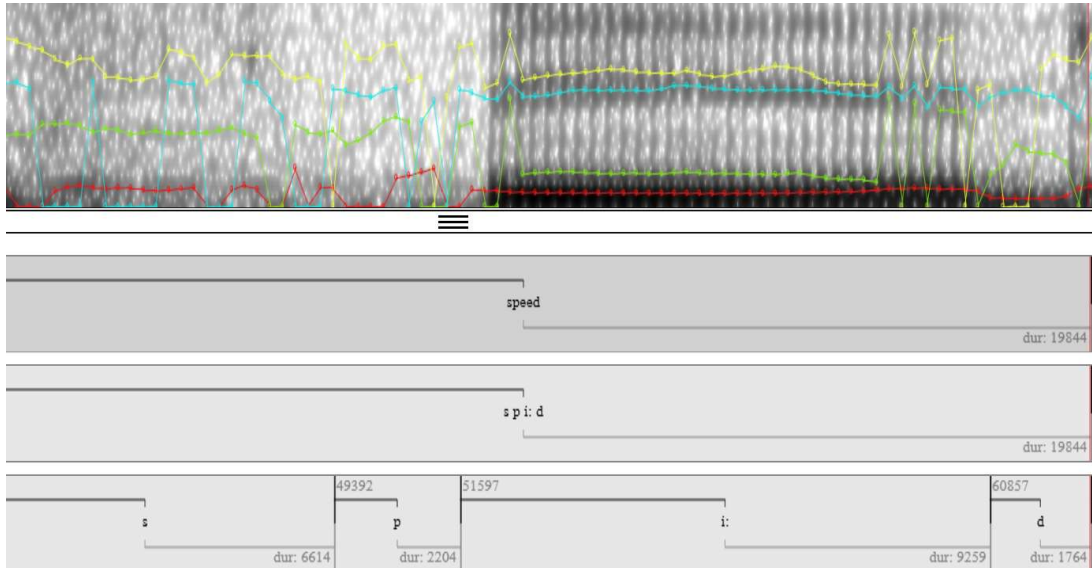
Figure 9. Spectrogram showing formant track error in the word 'speed': F3 (*blue*) automatically tracked along actual second formant; F2 (*green*) tracked along vowel harmonic rather than formant.

Of course there are some instances where formant track errors occur which cannot be fixed. Where formants cannot be identified as separate bands on a spectrogram, or through looking at a spectral slice they are excluded from the database, for instance Figure 8 above. Hand checking formants allows the exclusion of these tokens as they are discovered, rather than having to remove them at a later stage of analysis through more arduous means. Some further examples of formant errors and unusable vowels are given in Appendix B.

The second cause of formant calculation errors is related to speaker gender. Formant mistracking is far more common in recordings of female participants than male. This is because

formant analysis tools have largely been produced based on adult male speech, and do not take into account the higher formant frequencies produced by female speakers (Harrington, 2010). Adult females tend to have vocal tracts which are shorter than adult male vocal tracts, and this results in female formants being higher in frequency (Diehl, Lindblom, Hoemeke & Fahey, 1996). These gender effects mostly impact first formant tracking and back/low vowels, as F0 for high pitched female speech often falls between 150-450 Hz (Wang & Quatieri, 2010). This largely overlaps with typical F1 values for male speech, meaning that for high pitched female speech formant trackers will often mistake F0 for F1.

Certainly in the Auckland Voices dataset analyzed here, much more formant correction had to be undertaken for female speakers, than for male. This was especially true for noisy recordings, where background noise compounded with gender effects further impaired formant calculation. If formant correction had not been undertaken for these female speakers, an enormous number of tokens would have had to be excluded from analysis. Or else these tokens would have seriously compromised the accuracy of results. Again while the automatic calculation of formants saved valuable time, these formants cannot be relied on in their raw format.

4.1.4 Vowel Targets

This study extracts the F1 and F2 values of vowels at the "vowel target". The vowel target is a single time point in the duration of a monophthong (or two time points for a diphthong), which aims to capture the steady-state of a vowel's production (Harrington & Cassidy, 1999). It is the point of the vowel which is presumed to be least influenced by context effects (e.g. coarticulation). While vowel targets tend to fall around the temporal midpoint of a vowel, they do not have to. Using a "vowel target" is in opposition to simply extracting the F1 and F2 values at the vowel midpoint for monophthongs, or at 20% and 80% for diphthongs. This may or may not be the steady state of the vowel. Vowel targets are calculated for the data presented here based on criteria from Watson et al. (1998).

For the read speech analyzed, extracting vowels at the vowel target is particularly important. Extracting formant values at a vowel target reduces the impact of coarticulation on formant values; which improves the accuracy of the data. The read speech in this study by design contains large numbers of lateral consonants and other approximants. These consonants have particularly heavy coarticulation effects. Depending on their position, approximants can cause the vowel target to fall significantly earlier or later than the temporal midpoint of a vowel. If formant values were simply extracted at a midpoint many of the values would not fall during the steady state of vowel production, and the accuracy of our analysis would be weakened. In addition, some long vowels in New Zealand English, particularly high front FLEECE, are typically produced with a significant onglide. This results in the target or steady state of these vowels falling much later than the temporal midpoint.

The only major issue with using vowel targets in this study arose during diphthong labeling. Because vowels and therefore targets were labeled phonemically rather than phonetically, there were instances amongst the most variable speakers where a diphthong would require for instance a first target marked at a peak of F2, but the variant the speaker used would have an F2 trough rather than peak. Overall this did not have any major impact of the analysis, however.

4.1.5 Post Vocalic /l/

The reading passage performed in the Auckland Voices interviews by design, contains at least one instance of each NZE vowel preceding a lateral consonant. For many vowels there are multiple tokens preceding a lateral consonant. Coarticulation effects of post-vocalic /l/ are significant in New Zealand English. It is possible that the abundance of these tokens in read speech could impact the accuracy of an acoustic vowel analysis. This section considers the impact of lateral consonants on vowel formant values.

Lateral consonants are a type of approximant. Approximants share many features with vowels. Most importantly, vowels and approximants both belong to class of sounds known as

sonorants; or sounds produced with continuous non-turbulent airflow. Spectrally vowels and lateral consonants look very similar, and there is some debate as to whether the two sounds can or should be segmented given there is no discontinuity between the two (Harrington & Cassidy, 1999). In addition, there are many different allophones of these lateral consonants. Each allophone has varying coarticulation effects. The most common allophone, clear /l/, usually occurs at the syllable onset, and has the least impact on vowel formant values (Harrington & Cassidy, 1999). While coarticulation effects do occur, when clear /l/ precedes stressed vowels, those vowels will usually still reach the expected vowel target. Dark /l/, or velarized /l/, which usually appeared in syllable coda position, has a larger impact on formant values. Dark /l/ ([ɫ]) involves more tongue retraction and therefore has notably lower F2 frequency than clear /l/ (600-800Hz) (Sproat & Fujimura, 1993; Harrington & Cassidy, 1999). It also resists coarticulation influences of neighbouring vowels (Harrington & Cassidy, 1999). This means neighbouring vowels, especially front vowels, are more likely to have their F2 values reduced. Dark /l/ is very common in New Zealand English (Wells, 1982b; Maclagan, 1982; Johnson & Britain, 2007). In addition, speakers of New Zealand English increasingly use a vocalized version of /l/ in syllable coda position (Horvath & Horvath, 2001). This is a "vowelized" version of dark /l/, which has formant values similar to the New Zealand English THOUGHT vowel (Hardcastle & Barry, 1989). Its coarticulation effects are similar to those of dark /l/, but often more extreme. While this of course lowers F2 values of front vowels, it also has an extreme impact on FOOT. The FOOT vowel is particularly susceptible to coarticulation effects because of its short duration, and often assimilates entirely to the value of a following vocalized /l/. This is problematic for the read speech analyzed in this study, as there are very few tokens of FOOT which do not precede a vocalized /l/. As a result of the possible impact of post vocalic /l/ on the Auckland Voices read speech data, Section 5.3.1 investigates the acoustic impact of removing these tokens.

Chapter V

Method

5.1 Data

This research analyzes the vowels of young speakers from 3 different suburbs across Auckland. These are the West Auckland suburb Titirangi, the Central Auckland suburb Mount Roskill, and the South Auckland suburb Papatoetoe. These suburbs were selected for analysis based reasons outlined in Chapter 3, and briefly covered below. In addition, an older cohort of speakers from Titirangi were selected for analysis to consider whether there are any age effects between older and young Auckland speakers. Finally, to test whether there are differences between careful and casual speech, and consider some individual speakers, conversational speech from two groups: Papatoetoe men under 25 and Mount Roskill women under 25 was also prepared. In total over 12500 individual vowel tokens from 40 participants were prepared for analysis.

5.1.1 Data Selection

The acoustic study performed here considers read speech from young women and men under 25 from all three suburbs, as well as speech from a cohort of Titirangi women over age 40. The following section briefly discusses the sociolinguistic relevance of each suburb for this acoustic analysis.

Read Speech: Young Speakers

This thesis is interested in whether there is a relationship between diversity and sound change in Auckland, following Cheshire et al. (2011, 2013). If there is, then we expect to see evidence of this among our young speakers. Titirangi young speakers are expected to produce the most conservative vowel spaces relative to previous work on New Zealand English. There should also be less interspeaker variation in Titirangi, and our younger speakers are expected to sound

much the same as the older cohort also analyzed in this study. Papatoetoe young speakers' vowel spaces are expected to differ from New Zealand English more than Titirangi speakers', but we would expect less variation than the Mount Roskill group. This is because of the demographic makeup of Papatoetoe as discussed in section 3.2.2. "Pasifika English" features may also appear in this data set. Finally in the super diverse suburb, Mount Roskill, we expect evidence of linguistic diversity, and if there are changes from New Zealand English in Auckland, it is expected to be most extreme for these speakers. Because the diversification of Mount Roskill has occurred fairly recently, it is possible that this diversity will manifest as greater inter and intra-speaker variability i.e. speakers differing from one another and speakers producing multiple variants of the same phoneme. This often occurs in the pre-leveling stages of dialect formation (Trudgill et al., 2000). In addition we might expect some unusual combinations of linguistic features, in a similar vein to those found in the children of earliest migrants to New Zealand (Trudgill et al, 2000).

Read Speech: Older Speakers

This study largely focuses on the speech of young Aucklanders, and whether there is sound change occurring amongst these speakers on a cross suburb level. However, to do this there needs to be some sort of benchmark by which sound change can be measured. Comparison with previous acoustic vowel studies is one option, but different data collection and preparation methods make comparison between studies rather unreliable, especially in regards to statistical analysis. In addition to recordings with younger participants, however, the Auckland Voices project also contains recordings with older male and female participants from each suburb. Ideally this project would look at the vowel spaces of all speaker groups, young and old, male and female; but preparing such a large quantity of data is not feasible within the scope of this thesis.

Therefore, in addition to considering the results of previous studies, read speech from a group of older women from Titirangi has been prepared. Previous sociolinguistic work on New Zealand English suggests these women should be the most conservative speakers of New

Zealand English in the Auckland Voice Project. Maclagan, Gordon & Lewis (1999) suggest that while young women lead sound change in New Zealand, older women tend to be the most conservative speakers, especially in stable linguistic environments. Titirangi is the most homogenous suburb in the Auckland Voices database, which means it is likely to be the most linguistically stable (Cheshire et al., 2013). These two factors suggest the older Titirangi women will be our most conservative cohort of speakers. Given that we are predicting innovative changes in Auckland, and that past studies have suggested New Zealand English has no regional variation (Bauer, 1994; Gordon & Deverson, 1998), more conservative should entail more like traditional "New Zealand English".

Conversational Data Selection

As well as analyzing vowels from read speech for young male and female speakers from each suburb, conversational speech from two suburbs of interest has also been prepared and analyzed. While it would have been ideal to analyze conversational speech from all 7 speaker groups, there is a limit to how much data can feasibly be prepared within the scope of this project. With this in mind, conversational speech from interviews with Mount Roskill women, and Papatoetoe men were prepared for analysis.

The younger Mount Roskill females were selected as they should represent the most diverse speakers of New Zealand English in Auckland. Studies of NZE have suggested that it is young women who largely lead sound change (Maclagan et al., 1999; Maclagan & Hay, 2007; Gordon & Maclagan, 2001). In addition, linguistic diversity is also said to increase variation in large cities (Cheshire et al., 2013). The combination of these two circumstances suggests that Mount Roskill young women are likely to be the most innovative speakers in this data set. Read speech tends to be more conservative than conversational speech, so presumably there will be even greater variability in conversational speech. Therefore, analyzing conversational speech from of this group of young women might provide insight into the sociolinguistic factors shaping New Zealand English in Auckland.

As well as younger Mount Roskill women, a group of young male speakers from Papatoetoe have been selected for analysis. Many studies recently have tried to identify features of some kind of "Pasifika English". Papatoetoe is well known for having a large "Pasifika" community, and this community is well represented in our data set. "Pasifika English" is difficult to identify because speakers tend to revert to more conservative NZE forms in the formal interview settings that are usually involved in linguistic data collection. If this is true, we might expect Pasifika English features to be present in the conversational speech of this group of speakers, but not in read speech. In order to test this theory, the analysis of at least one group of speakers from this suburb was necessary. The younger Papatoetoe men were selected over the younger Papatoetoe women for primarily pragmatic reasons. First, it was not feasible in the time frame available to prepare data for all speakers from this suburb, and a larger group of male speakers were available for analysis. Second, analysis of the reading passage suggested that the recordings of the young men tended to be of better quality than the young women. Third, because of speech software design men's speech tended to require less manual formant correction than the speech of women.

5.2 Data Preparation

Data preparation plays a particularly important role in this study. This section outlines the rigorous process undertaken to transform the raw data into the format necessary for acoustic analysis. The care undertaken in data preparation makes vowel analysis results more accurate and therefore conclusions more reliable. The importance of careful data preparation techniques in discussed in detail in Chapter 4. While the preparation of the read speech and continuous speech was similar in many ways, the two styles necessitated some key differences in preparation methods. For this reason each process is discussed separately.

Data was taken from interviews conducted as a part of the Auckland Voices Project discussed in Chapter 3. Speech was recorded on two devices: a Marantz PMD 661 and a Zoom H5 using TDK lavalier clip-on microphone. The speech signal was sampled at 44.1 kHz and quantized to 24 bit.

In two instances where the main microphone failed, backup recordings from the Zoom H2 were used. For these recordings the speech signal was also sampled at 44.1 kHz, but quantized to 16 bit.

5.2.1 Read Speech

The number of reading passages analyzed for each suburb/gender is given in Table 1. A total of 39 passages of read speech were prepared and analyzed. As well as male and female younger speakers from each suburb, a group of older speakers from the suburb of Titirangi were selected for analysis. This group was selected as they would, for sociolinguistic reasons outlined in section 5.1.1, likely represent a more conservative NZE dialect than any other speaker group available. Ideally these speakers would provide a benchmark with which to compare younger speakers when discussing sound changes.

Table 1.
Speaker distribution by suburb and gender

Gender	Titirangi	Papatoetoe	Mount Roskill	Titirangi - Older
Females	3	6	8	7
Males	3	7	6	-

Unfortunately the number of young participants analyzed from Titirangi is considerably smaller than the numbers for the other two suburbs. The read speech of four other participants was prepared, but ultimately had to be excluded. The excluded participants were all aged between 7 and 12. As a result of their young age, these participants' voices had not lowered yet. This meant their formant values were simply not comparable or combinable with the other speakers in the database.

The process of preparing the reading passage for analysis in R is as follows: First the reading passage had to be located and cropped from the larger interview .wav file. The read speech ranged from 2-4 minutes in length, depending on speech rate. A smaller ELAN transcription file

for each reading passage then was created with text copied from the interview transcription file. The transcription file for each read passage contains 33 utterances; each corresponding to an equivalent utterance from the read passage in *Appendix A*. This organization made database searches easier later in analysis. From this ELAN file individual .txt files were created for each utterance, and matching .wav files were cropped using *Audacity(R) recording and editing software, version 2.2.0* (Audacity Team, 2017).

Forced alignment was then performed using the WebMAUS software, English (NZ) service (Kisler, Reichel, & Schiel, 2017). This provided an automatic phonetic transcription of the text in *Praat TextGrid* format. All phoneme boundaries were then hand checked and corrected in *Praat* (Boersma, 2001)*,* and prosodically stressed vowels were marked on a separate tier*.* For further discussion on the necessity of hand-checking phoneme boundaries, and the selection of stressed vowels for analysis see Sections 4.1.2 and 4.1.1 respectively.

After hand correction in PRAAT, TextGrids were converted to EMU-Web-App (Winkelmann & Raess, 2014) "bundles", to calculate formants. The formant tracks of all stressed vowel tokens were then hand checked in the EMU-Web-App. Further discussion on formant correction is given in Section 4.1.3. The *vowel target(s)* of each prosodically stressed monophthong and diphthong were then marked on a separate tier in the EMU-Web-App, according to criteria from Watson et al. (1998); one target for monophthongs and two for diphthongs. For instance, the targets of close front vowels were marked at the peak of F2, while the targets of open vowels were marked at the peak of F1.

Finally the first and second formants were extracted at the vowel target for monophthongs, and formant tracks between the first and second targets were extracted for diphthongs in `R` (R Core Team, 2013), using `EmuR` (Winkelmann, Jaensch, Cassidy & Harrington, 2018). For further discussion on the extraction of formant values at the vowel target as opposed to vowel "midpoint" see Section 4.4.1. Further formant analysis was then undertaken in `R` using `EmuR`.

5.2.2 Conversational speech

Conversational speech from interviews with Mount Roskill women (n=6), and Papatoetoe men (n=7) was also prepared for analysis. Data preparation for continuous speech resulted in data in the same format as the read speech. There were some differences in workflow due to differences in the format of the casual and read speech, however. Given that interviews ranged from 45 minutes to 1 hour 30 minutes, it was decided that 10 minutes of speech would be analyzed from each interview. This served to limit data preparation time, and ensured that speakers with longer interviews were not over-represented in the data set. An initial test also indicated 10 minutes would provide an adequate number of vowel tokens for analysis. First, 10 minutes of speech was cropped from the 30 minute mark of each interview. This ensured that the speaker was comfortable, i.e. not too nervous, as they may be towards the beginning of an interview; or fatigued towards the end of an interview. Both of which may impact speech production.

This speech then had to be chunked into utterance length .wav files in *Audacity* (Audacity Team, 2017), and assigned a corresponding .txt file copied from the ELAN (Sloetjes & Wittenburg, 2008) transcription file. Although there are some automatic services for chunking data available, it was easier for multiple reasons to do this manually. First, it was convenient to be able to manually crop out the speech of the interviewer at this stage. Second, some filler words and pauses create problems when forced alignment is performed, making hand correction a more arduous task. This is further discussed in section 4.1.2. Manually chunking the data allowed these problematic segments to be removed prior to forced alignment, reducing future workload. Finally, transcription errors were able to be fixed and corrected during hand chunking, rather than requiring phoneme-by-phoneme editing at a later stage. These 10 minute segments provided between 75 and 160 utterances for analysis per speaker, depending on participants willingness to communicate.

As with the read speech, forced alignment was then performed using the *WebMAUS* software, English (NZ) service (Kisler, et al., 2017). This produced *Praat TextGrids* (Boersma, 2001)*.* Rather

49

than correcting phoneme boundaries and marking prosodically stressed vowels in *Praat* as with the reading passage, these TextGrids were immediately converted to *EMU-Web-App* (Winkelmann & Raess, 2014) "bundles". This allowed streamlining of the data preparation process. Formants were then calculated using `Forest` in R (R Core Team, 2013), using `EmuR` (Winkelmann, Jaensch, Cassidy & Harrington, 2018). Converting the files to EMU-Web-App format prior to correcting phoneme boundaries and marking stress meant that data preparation could be accelerated. The correction of prosodically stressed vowels, correction of formants where necessary, and marking of vowel targets on a separate tier could be completed at the same time; rather than undertaking each process separately. Stressed syllables also did not require labeling on a separate tier as the same effect was achieved by only marking the vowel targets of stressed vowels. This was only possible for the continuous speech due to knowledge gained while preparing the read speech. Finally formants were extracted at the vowel target for monophthongs, and formant tracks between the first and second targets were extracted for diphthongs, in R (R Core Team, 2013), using `EmuR` (Winkelmann, Jaensch, Cassidy & Harrington, 2018).

5.3 Data Analysis

Following data preparation, data analysis was undertaken in R largely using `EmuR`. This includes plotting monophthongs and diphthongs on an F1/F2 plane, calculating vowel duration for all vowels, and onglide for relevant monophthongs, as well as performing statistical analysis on read speech. This section gives total vowel counts for both diphthongs and monophthongs, and information about some necessary token removal at this stage. The total number of monophthongs considered in this study is given in Table 2 below.

Table 2.
No. Monophthong Tokens by Suburb & Gender

Suburb	Read Speech		Conversational speech	
	Male	*Female*	*Male*	*Female*
Papatoetoe	792	626	2252	
Mount Roskill	600	867		1868
Titirangi	309	337		
Titirangi - Older		783		
			Total Monophthongs:	8434

The total number of diphthongs considered in this study in given in Table 3 below.

Table 3.
No. Diphthong Tokens by Suburb & Gender

Suburb	Read Speech		Conversational speech	
	Male	*Female*	*Male*	*Female*
Papatoetoe	450	402	818	
Mount Roskill	327	525		730
Titirangi	187	196		
Titirangi - Older		486		
			Total Diphthongs:	4121

5.3.1 Post Vocalic /l/ Revisited

At this stage it was decided that all vowel tokens in syllables with a lateral in coda position would be removed from analysis. This was because the reading passage by design contained every vowel preceding a lateral. This resulted in vowel plots being skewed by the heavy coarticulation effects of this vowel. The impact of removing these tokens is demonstrated in Figure 10 below, where tokens with an /l/ in coda position are consistently retracted. The impact of post-vocalic /l/ on each database can be found in Appendix C. Removing these tokens obviously impacted monophthong token numbers. For in depth information about how removing post vocalic /l/ impacted token counts, see Appendix D.

51
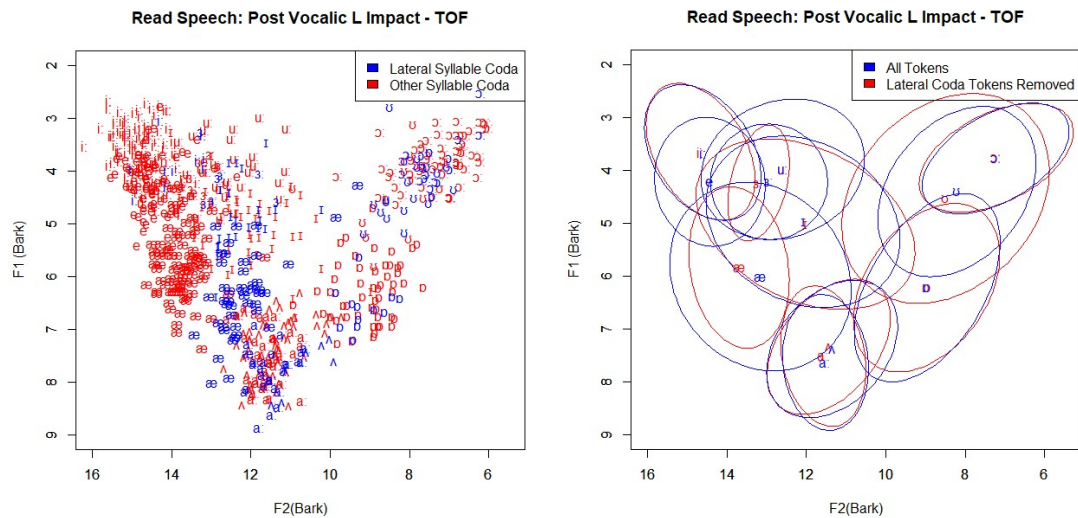
Figure 10. Effects of post vocalic /l/ on individual vowel tokens (*left*), and centroids (*right*) - Older Titirangi Speakers.

Because post vocalic /l/ is not abundant by design in the conversational speech analyzed, and because this data is not being compared statistically or visually with the read speech, post-vocalic /l/ tokens have not been removed from the analysis of conversational speech.

Tokens preceding an /l/ were also not removed from the a diphthong analysis. A comparison of the control group of older Titirangi speakers with previous studies of NZE diphthongs such as Watson et al. (1998) suggested that the impact of post vocalic /l/ was not large enough to warrant excluding these tokens. Furthermore, visual comparison of the conversational diphthongs and read speech diphthongs further confirmed that the impact of post-vocalic /l/ on these tokens was minimal. Diphthong plots require averaging of formant tracks, whereas monophthong plots involve plotting an ellipse with a 95% confidence interval, which is where post-vocalic /l/ tokens are most problematic. It should be noted that where post vocalic /l/ did occur, second target values were impacted to varying degrees. They did not however, impact analysis enough to warrant their exclusion.

5.4 Statistics

In addition to the acoustic analysis, two statistical analyses were performed on the read speech data. The first considers gender and suburb effects, comparing young male and female speakers from each suburb. The second considers age effects and compares younger and older speaker groups. For both analyses linear mixed effects models were built and compared. As a result of differences between male and female speech, the comparison of young speakers also involved a data transformation to make frequency based observations more comparable. The following sections describe this transformation as well as the statistical modeling process.

5.4.1 Data Transformation

Male and female vowel spaces are usually compared visually (Watson et al, 2016; Maclagan & Hay, 2007; Easton & Bauer, 2000). It is difficult to compare them empirically due to the impact the gender differences in vocal tract length have on the formants (Harrington, 2010; Diehl, Lindblom et al., 1996). For the statistical analysis described in Section 5.4.2 it is necessary to be able to compare male and female speakers, however. It is possible to make male and female data comparable by performing a linear transformation. This transformation can convert male speaker values closer to female speaker values, or female speaker values closer to male speaker values. In this study, because of the greater number of female speakers, a male to female transformation has been performed. This transformation is shown in the equation below.

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad (1)$$

$F_1;F_2$ are female speaker formant values and $M_1;M_2$ are male speaker formant values. $\mu_1, \mu_2, A_{11}; A_{12}; A_{21}; A_{22}$ are a set of six numerical constants. The calibration of this transformation requires the selection of 3 points in the male a distributions and the corresponding three points in the female distributions onto which the male distribution will be mapped.

$$\left(\begin{bmatrix} F_{11} \\ F_{12} \end{bmatrix}, \begin{bmatrix} M_{11} \\ M_{12} \end{bmatrix}\right), \left(\begin{bmatrix} F_{21} \\ F_{22} \end{bmatrix}, \begin{bmatrix} M_{21} \\ M_{22} \end{bmatrix}\right), \left(\begin{bmatrix} F_{31} \\ F_{32} \end{bmatrix}, \begin{bmatrix} M_{31} \\ M_{32} \end{bmatrix}\right) \qquad\qquad (2)$$

M_{ij}, and F_{ij} are the centroids of the point vowels /iː, aː, ɔː/. Using these values the six unknown constants can be found, and the data can be transformed using the equation in (1). This transformation essentially involves rotating and stretching the male data plane so that the three anchoring point vowels /iː, aː, ɔː/ match. Figures 11 and 12 below show the effects of transforming the male data space to match the females for the young Auckland voices speakers, and the impact this transformation has on the monophthong centroids.
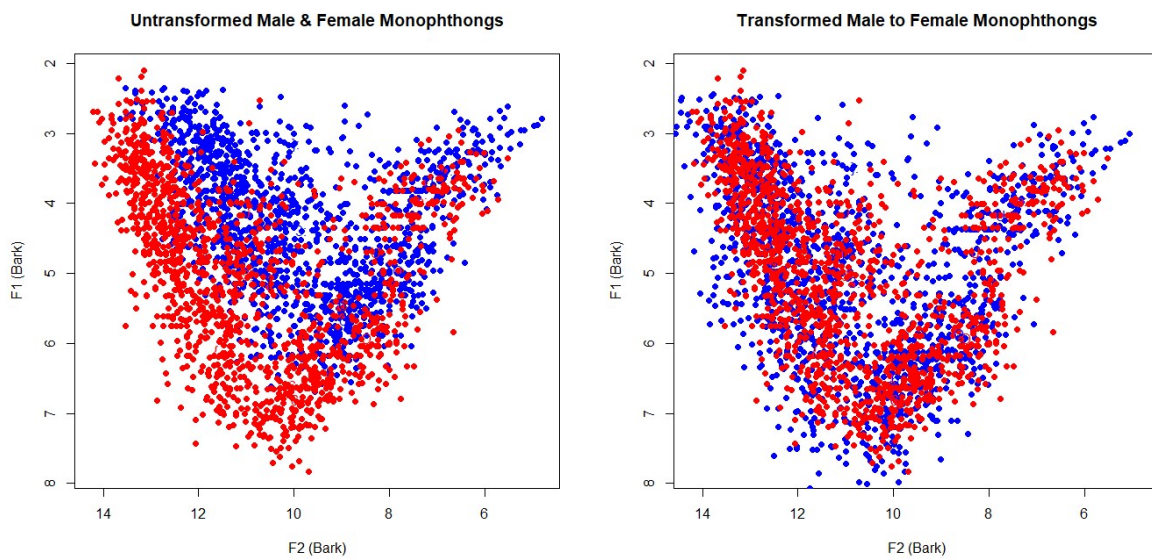
Figure 11. Raw Male and Female data (*left*), Transformed Male Data (*right*). This shows the impact of data transformation performed for comparing male and female speakers.
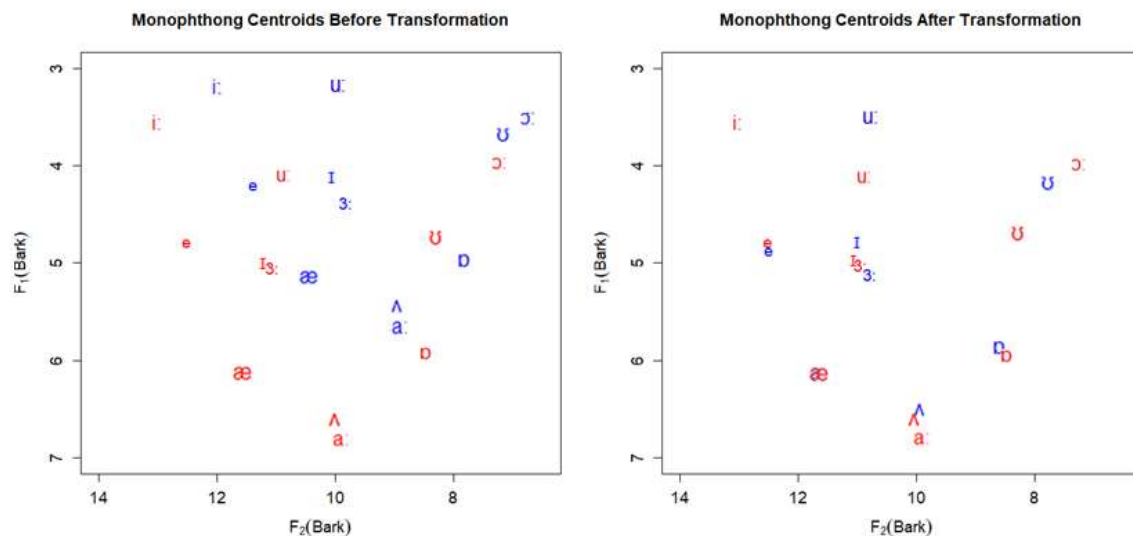
Figure 12. Vowel centroids before transformation (*left*), and after transformation (*right*). This shows the impact of data transformation performed for comparing male and female speakers on vowel centroids.

5.4.2 Statistical analysis

In this section the statistical analysis performed in Section 6.1.2 is described. This analysis uses linear mixed effects models to measure whether there are significant differences between data sets pertaining to the research questions in Section 3.3. For young speakers, gender, suburb, and interactions between gender and suburb are considered. An additional analysis comparing older and younger female speakers is also performed. The only difference between the two analyses is that the all female age effects analysis did not require vowel space transformation. This statistical analysis models F1 and F2 values simultaneously in the 2-D formant space. This provides additional statistical power to measure changes in any direction on the formant plane. This is in contrast to typical analyses which treat F1 and F2 separately (for instance, Watson et al, 2016; Maclagan & Hay, 2007; Easton & Bauer, 2000; Warren 2017).

All statistical analysis was done in R (R Core Team, 2013), and the linear mixed models were performed using the `lme` function in the `nlme` package (Pinheiro, Bates, DebRoy, Sarkar & R Core Team, 2018). The linear mixed effects models are compared using the `anova` function

also in the `nlme` package. Both the cross suburb and age analyses consisted of comparing pairs of linear mixed effects models for each New Zealand English vowel, for observations of the first and second formant. Both models treat Speaker as a Random effect. For the cross suburb analyses of young speakers the fixed effects were as follows: In the first model fixed effects were formant type (F1 or F2), and in the second model the fixed effects were Gender, Suburb and formant type (F1 or F2). For the analyses of age effects the fixed effects were: In the first model formant type (F1 or F2), and in the second model Age and formant type (F1 or F2).

Chapter VI
Results

This chapter presents the acoustic analysis results. The results are separated into four sections. The first section focuses on monophthongs; presenting vowel plots followed by a statistical analysis (6.1). The second section (6.2) presents diphthongs; and the third (Section 6.3) considers vowel length for all diphthongs and monophthongs. Finally section 6.4 calculates the vowel onglide for 3 relevant monophthongs: FLEECE, DRESS and GOOSE.

Male and female data is largely considered separately in the following sections. Speaker normalization also has not been performed on the acoustic vowel plots presented. Anatomical differences between speakers results in each speaker having a slightly different shaped vocal tract, which produces a slightly different vowel space. Normalization can reduce these speaker effects, but can also distort the data in erratic ways (Disner, 1980). The statistical analyses in Section 6.1.2, however, does combine male and female speakers, this is enabled due to the data transformation process outlined in Section 5.4.

Speaker Rhotacization

Before presenting the vowel analysis it is important to note that some speakers unexpectedly use rhotic dialects. Three speakers from Mount Roskill, two female and one male, consistently use a rhotic dialect throughout conversational and read speech. This accent is unlikely to be related to the Southland rhotic dialect found in New Zealand, and is likely acquired from parents or caregivers who do not speak New Zealand English. Most Papatoetoe speakers use a semi-rhotic dialect, where /ɹ/ is pronounced usually following NURSE, but also in words such as "year", "four" and "more". This rhotacization is not restricted to stressed vowel contexts; for instance the final /ɹ/ in the word "older" is often pronounced. This is a feature commonly associated with "Pasifika" New Zealand English dialects (Gibson & Bell, 2010).

6.1 Results of Monophthong Analysis

This section first compares the main monophthong data performed in read speech, and then examines the smaller selection of conversational speech. For the read speech, two primary variables and one secondary variable have been considered. The primary variables considered are suburb and gender across all younger speakers. These, for reasons given in Section 5.1.1, are compared with a single cohort of older speakers from Titirangi. The older speakers are considered first to establish their similarity to previous studies on New Zealand English. This is followed by centroids for 3 suburbs separated by gender. Continuous speech is then presented separately and is intended to provide a narrow analysis of two subgroups.

6.1.1 Monophthong Plots: Read Speech

This section presents ellipse plots in the F1/F2 plane for all monophthongs analyzed. It should be noted that for each group FOOT, GOOSE, and NURSE have been plotted as centroids without ellipses. This is because there are insufficient tokens of these vowels in the read passage to provide a meaningful ellipse. These vowels are considered in conversational speech in Section 6.1.3. The centroids have been included to indicate where these vowels might fall, but should only be considered a guide. For token counts of individual vowels see Appendix D. The observations gleaned from the ellipse plots are supported by statistical analysis results in Section 6.1.2. All vowel spaces presented in this section involve non-transformed data. The data transformation outlined in section 5.4.1 is used exclusively for the statistical analysis.

Figure 13 shows the ellipse plots in the F1/F2 plane for the monophthongs for the older Titirangi female speakers. This is compared with data from Watson et al. (1998). Visually these plots look strikingly similar, although because of methodological differences they cannot be compared statistically. In Figure 13 below (*left*), DRESS and TRAP are raised, a feature of New Zealand English discussed in Warren, 2017; Bauer, 1979; 1992 and Trudgill, Gordon & Lewis, 1998. There is a large amount of overlap between DRESS and FLEECE, supporting Warren's analysis that DRESS is rapidly approaching FLEECE. As expected KIT is retracted and lowered. START and STRUT form a long short pair. NURSE is raised, as is THOUGHT. Overall analysis of these speakers

produces a vowel space that is unremarkable in relation to previous studies on New Zealand English.
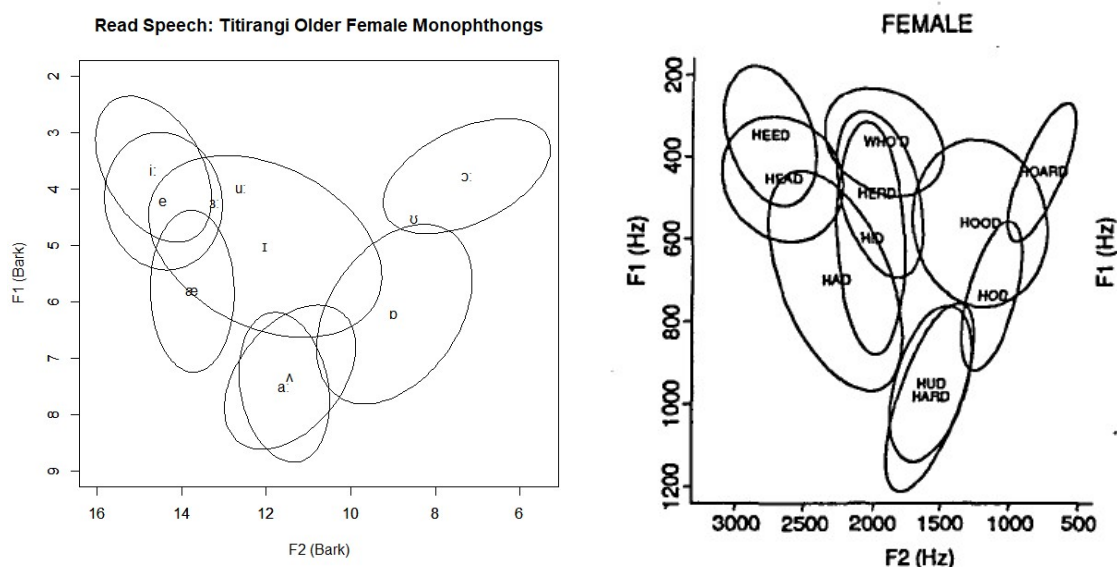
Figure 13. Ellipse plot presenting vowel centroids in the F1/F2 space of older Titirangi women from the AKL voices database *(left)*, and female New Zealand English speakers from Watson et al., 1998 *(right)*. The centroid of each vowel is marked by the appropriate IPA as outlined in chapter (1) *(left)* or with /hVd/ words *(right)*. The IPA symbols *(left)* are phonemic representations rather than phonetic. The boundary of each ellipse indicates a 95% confidence interval.

Figure 14 below shows ellipse plots in the F1/F2 plane for the monophthongs for young Titirangi speakers. There are no visible gender differences, and this is confirmed by the statistical analysis in Section 6.1.2. There are clear differences between the two younger groups and the older cohort, however. Most notably TRAP and DRESS are significantly lowered. NURSE is also lowered, but low token numbers for all groups makes an analysis of this vowel unreliable. Given the homogeneity of Titirangi insinuated that there should be little difference between older and young speakers, this result is surprising. Some features of New Zealand English do not appear to differ between older and young speakers, however. KIT remains retracted and

59

lowered, THOUGHT is raised, and START and STRUT form a long short pair. Ellipses suggest large amounts of inter or intra speaker variation amongst all speakers for TRAP, and DRESS and LOT for male speakers. Particularly for the front vowels which suggests a change in progress.
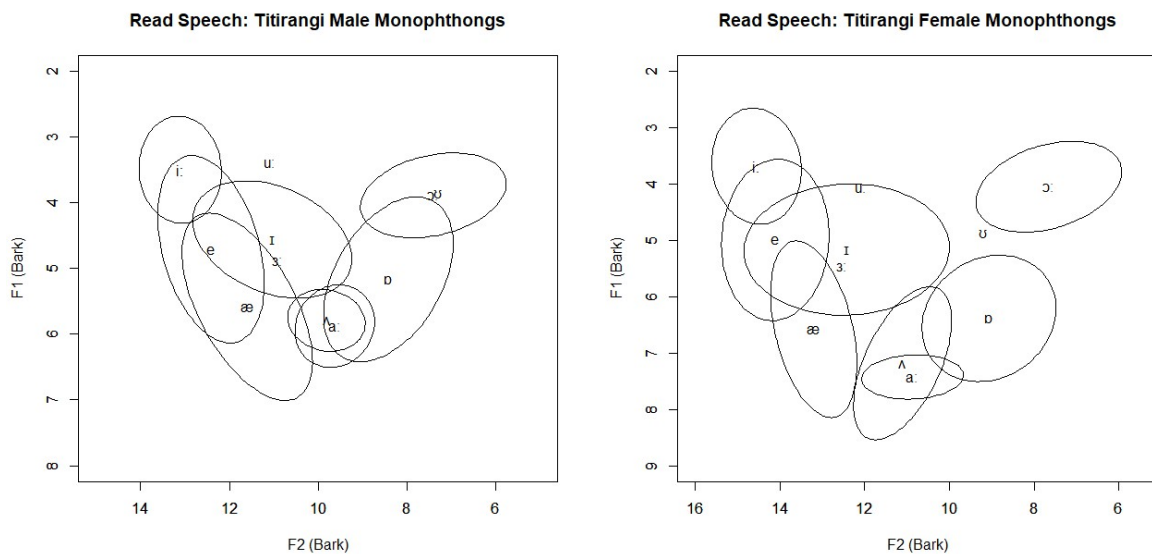
Figure 14. Ellipse plots in F1/F2 space presenting vowel centroids of younger speakers from Titirangi, male (*left*) and female (*right*) speakers. The boundary of each ellipse indicates a 95% confidence interval.

Figure 15 below shows ellipse plots in the F1/F2 plane for the monophthongs of young Mount Roskill speakers. The statistical analysis suggests some small differences between male and female speakers, for instance for male speakers, a slightly raised KIT, and slightly centralized THOUGHT, which can be seen on the plots above. More detail on statistical differences between genders is given in section 6.1.2.1. The Mount Roskill speakers differ from the older cohort in the same way as the young Titirangi speakers. TRAP and DRESS are considerably lowered, as is NURSE but again low token numbers make analysis of nurse unreliable.

60
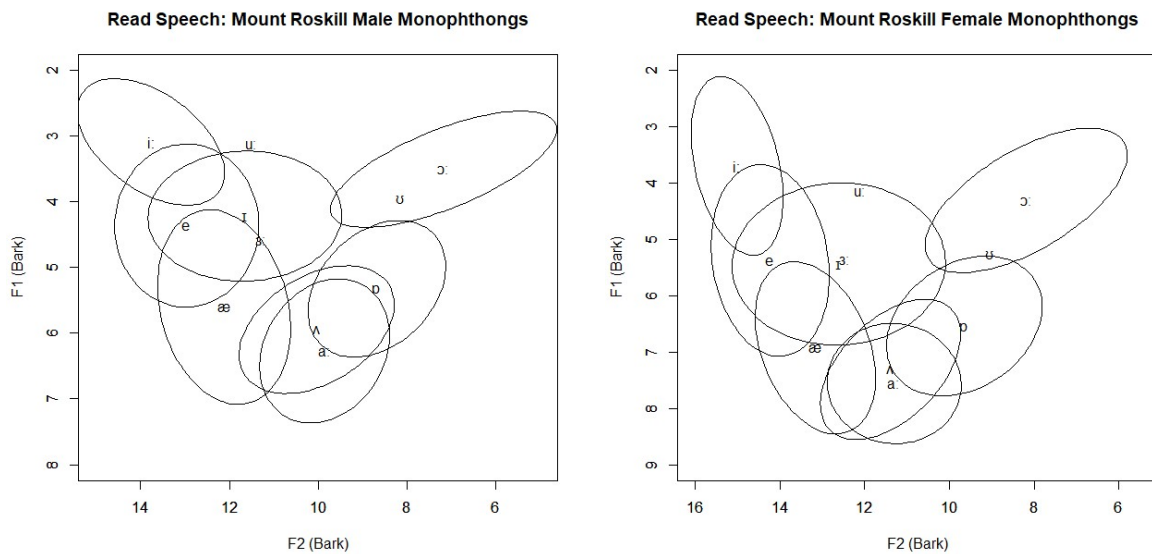
Figure 15. Ellipse plots in F1/F2 space showing vowel centroids of younger speakers from Mount Roskill, male (*left*) and female (*right*) speakers. The boundary of each ellipse indicates a 95% confidence interval.

As with the Titirangi cohorts in Figure 14, some features of New Zealand English do not appear to differ between young Mount Roskill speakers and the older Titirangi group. KIT remains retracted and lowered, THOUGHT is raised, and START and STRUT form a long short pair. LOT also does not differ from the older group. There is also evidence of interspeaker or intraspeaker variation within this group. Ellipses for TRAP suggest large height variation for both groups, and variation in DRESS height for young women. There also seems to be some centralization of THOUGHT. KIT has a large ellipse for both groups, but this is possibly because of coarticulation effects related to its short duration.

Finally, Figure 16 below shows the ellipse plots in the F1/F2 plane for the monophthongs of young Papatoetoe speakers. The statistical analysis in Section 6.1.2 showed no significant differences between male and female speakers. The vowel spaces are also much the same as the young speakers from Mount Roskill and Titirangi above. DRESS and TRAP are lowered, as is NURSE. KIT, LOT, START and STRUT remain in their traditional NZE position.

61
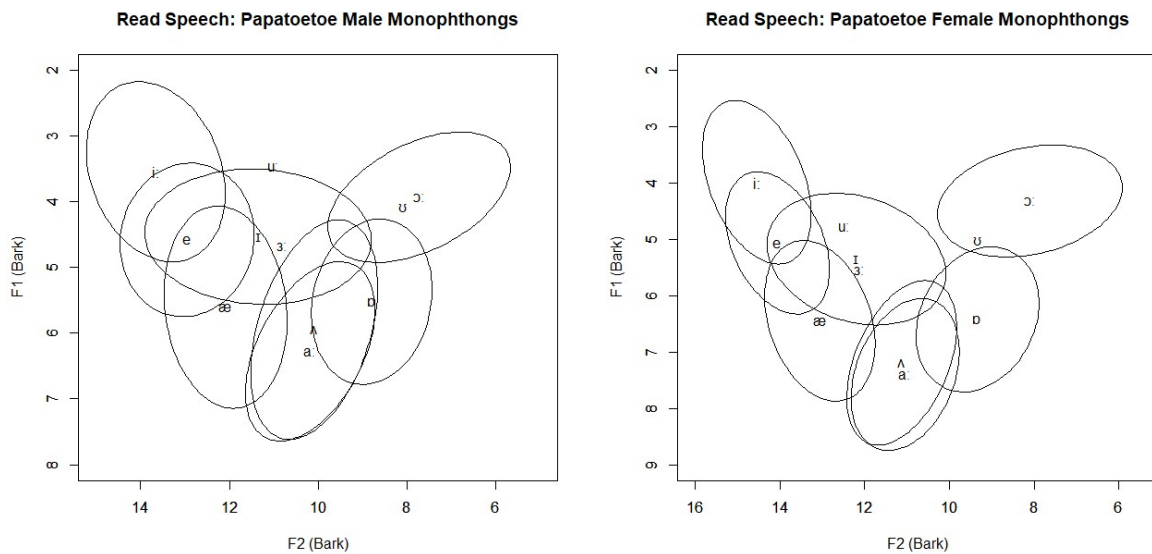
Figure 16. Ellipse plots in F1/F2 space showing vowel centroids of younger speakers from Papatoetoe, male (*left*) and female (*right*) speakers. The boundary of each ellipse indicates a 95% confidence interval.

The statistical analysis found some differences between Papatoetoe speakers and Young Mount Roskill women, but these will be discussed in the following section. Particularly for the young men, there appears to be large height variation for the low vowels START and STRUT. This is most likely due to speaker differences, not actual variation. One speaker in particular had a very compact vowel space, and because data was not normalized this resulted in the height variation seen here.

Overall largest monophthong differences are between older and younger speakers. There are some slight differences between young groups, but these are not visually apparent, and are only revealed through statistical analysis below. The vowels undergoing change amongst young Auckland Speakers are DRESS, TRAP and possibly NURSE, although there are not enough tokens to reliably analyze NURSE. Given that TRAP and DRESS appear to be lowering, it is interesting that KIT is not raising indicating a reversal of the chain shift commonly discussed in New Zealand English literature. In fact the statistical analysis below indicates that young female speakers actually realize KIT significantly lower and further back that the older cohort. It is possible that this is

because KIT height often correlates with DRESS height in New Zealand English, and the older group has a raised DRESS and therefore slightly raised KIT. Some vowels appear to be stable across both older and young speakers. These are LOT, FLEECE, START and STRUT. The statistical analysis below explores the above observations. It confirms that there are some significant differences between older and younger groups. It also explores some differences between the young Mount Roskill women and other young groups, supporting the hypothesis that this group is the most variable.

6.1.2 Results of Statistical Analysis

This section presents the results of the statistical analyses described in section 6.1.1. Results of the cross suburb/gender analysis are presented first, followed by the analysis comparing younger and older female speakers. FOOT has been removed from this analysis because of low token numbers. For token numbers by vowel, see Appendix D.

When significant differences were identified in the ANOVAs of the pairs of models, t-tests were used to investigate where the significant differences lay amongst the various fixed effects. In this analysis suburb or gender was coded as a factor, with "Mount Roskill" as the suburb reference level, and "female" as the gender reference level. Both formants are listed as factors, but the first formant was the reference level. As a result of this structure, full t-test results are required to interpret results. For space reasons these are not presented in-text, but relevant tables can be found in Appendix E.

6.1.2.1 Suburb and Gender Effects

Table 4 below presents results from specific vowel analyses that returned significant effects when comparing the null model (fixed effects formant type) with the model based on formant and suburb (fixed effect formant type and suburb). This determines whether there are any statistical differences between suburbs.

Table 4.
*Null model (type-1) vs (type-1)*sub*

	Degrees of Freedom	AIC Difference	Log Likelihood Ratio	P value
FLEECE	12	4.926	12.926	0.04
NURSE	12	1.9943	9.994	0.012

Further analysis of t-tests, reveals more detailed information about these significant differences. For FLEECE, significant differences in F1 occur between Mt Roskill and Papatoetoe. Significant differences in F2 are between Mount Roskill and Papatoetoe and Mount Roskill and Titirangi. For NURSE, significant differences are between Mount Roskill and Titirangi in F1 and Mount Roskill and Papatoetoe in F2. These significant differences are presented in Table 5 below. Full t-test results for significant vowels are presented in Appendix E.

Table 5.
*Null model (type-1) vs (type-1)*sub – t-tests.*

		Value	Std.Error	DF	t-value	p-value
FLEECE	typeF1	3.38172	0.08744	1054	38.67543	<.0001
	typeF2	13.23940	0.10787	1054	122.73907	0
	subP	0.35459	0.12583	31	2.81808	0.00834
	typeF2:subP	-0.60613	0.23805	1054	-2.54619	0.01103
	typeF2:subT	-0.76763	0.30146	1054	-2.54635	0.01103
NURSE	typeF1	4.94378	0.08407	85	58.80371	<.0001
	typeF2	11.04592	0.14619	85	75.55829	<.0001
	subT	0.34101	0.14489	30	2.35353	0.02534
	typeF2:subP	-0.58158	0.26317	85	-2.20985	0.029803276

Note. subP = Papatoetoe, subT= Titirangi

Following the cross suburb comparison, the null model (fixed effects formant type) is compared with the model based on formants and gender (fixed effect formant type and suburb). This determines whether there are any gender based differences in the dataset. The only significant gender difference was for GOOSE. Further analysis reveals that there is a significant height

difference (F1) between male and female speakers; as shown in Tables 6 and 7 below. The F1 of male speakers is slightly lower than the F1 of female speakers. However, tokens of GOOSE are low, so this result is unreliable.

Table 6.
Null model vs (type-1)*sex

	Degrees of Freedom	AIC Difference	Log Likelihood Ratio	p-value
GOOSE	10	8.0195	12.01354	0.0025

Table 7.
Null model vs (type-1)*sex – t-test

	Value	Std.Error	DF	t-value	p-value
typeF1	4.084	0.09966	203	40.98114	<.0001
typeF2	10.8763	0.17234	203	63.11	<.0001
sexM	-0.56564	0.1459	32	-3.8769	0.00049

Note. sexM = male

The final analysis compares the null (fixed effects formant type) model to a model based on formants, suburb and gender (fixed effects formant type, suburb and gender). This revealed significant interaction between type, suburb and gender for TRAP, FLEECE, KIT, THOUGHT and GOOSE as shown in Table 8 below.

Table 8.
Null model vs. (type-1)*sub*sex

	Degrees of Freedom	AIC Difference	Log Likelihood Ratio	P value
TRAP	18	2.301	22.3	0.0136
FLEECE	18	0.146	20.146	0.0279
GOOSE	18	6.6485	26.649	0.003
THOUGHT	18	0.4216	20.421	0.0255
KIT	18	8.6464	28.446	0.0015

The significant interactions between suburb and gender are presented in Table 9 below. For each vowel the null model F1 and F2 results are presented as well as any significant interactions. Further analysis of ᴛʀᴀᴘ reveals that the interaction is minimal, and only involves a slightly lower F1 and higher F2 value for young male Mount Roskill speakers compared to the Mount Roskill Women, as well as a higher F1 for Papatoetoe women compared to the Mont Roskill women. This is not immediately evident when visually comparing the two vowel spaces in Figures 15 *(males left, females right)* and 16 *(right)* from the previous section.

Table 9.
Null model vs. (type-1)*sub*sex – t-tests

		Value	Std.Error	DF	t-value	p-value
TRAP	typeF1	3.8977	0.11792	199	33.05326	<.0001
	typeF2	10.7879	0.25816	199	41.78773	<.0001
	subP*	0.5276	0.17201	28	3.06788	0.00475
	sexM	-0.58652	0.19015	28	-3.0846	0.00455
	typeF2:sexM	1.00578	0.43678	199	2.30271	0.02233
FLEECE	typeF1	3.4870	0.1102	1051	31.6525	5.64E-155
	typeF2	13.2773	0.1359	1051	97.7137	0
	typeF2:subP*	-0.7169	0.3205	1051	-2.2368	0.0255077
GOOSE	typeF1	3.8977	0.1179	199	33.0533	<.0001
	typeF2	10.7879	0.2582	199	41.7877	<.0001
	subP†	0.5277	0.1720	28	3.0679	0.0047457
	sexM†	-0.5865	0.1901	28	-3.0846	0.0045517
	typeF2:sexM*	1.0058	0.4368	199	2.3027	0.022329
THOUGHT	typeF1	4.0085	0.0695	519	57.6656	<.0001
	typeF2	7.4127	0.1480	519	50.0777	<.0001
	sexM†	-0.3284	0.1066	28	-3.0804	0.0046001
	subP:sexM†	0.4800	0.1533	28	3.1313	0.0040492
	subT:sexM†	0.6106	0.1926	28	3.1708	0.0036662
KIT	typeF1	5.0103	0.0747	417	67.0890	<.0001
	typeF2	11.1779	0.1132	417	98.7864	<.0001
	sexM†	-0.5089	0.1225	28	-4.1542	0.0002775
	typeF2:sexM†	0.6535	0.2274	417	2.8736	0.0042655
	subP:sexM*	0.4499	0.1704	28	2.6405	0.0133845
	subT:sexM*	0.5394	0.2048	28	2.6336	0.0136037
	typeF2:subT:sexM†	-1.0096	0.3727	417	-2.7089	0.0070283

Note. T= Titirangi, P= Papatoetoe, M= male, No suburb label= Mount Roskill, No gender label = female.

For FLEECE the only significant difference is between Young Mount Roskill and Papatoetoe women. The Papatoetoe young women have slightly retracted FLEECE. It is unclear whether this difference simply reflects anatomical differences between speakers, or is a legitimate difference. For GOOSE there are significant F1 differences between female Mount Roskill and Papatoetoe speakers, as well as both F1 and F2 differences between Mount Roskill men and women. This is evident in Figures 15 and 16. Papatoetoe women have a significant lowered GOOSE vowel compared to the Mount Roskill women, while the male Mount Roskill speakers GOOSE is raised and fronted compared to the Mount Roskill women. Results for GOOSE are unreliable because of low token numbers, however.

As for THOUGHT, Mount Roskill female F1 values are significantly different from all three male groups. Male speakers' THOUGHT vowels, for all suburbs, are lowered compared to the Mount Roskill women; with the greatest difference being the Papatoetoe men. This is visible on the vowel spaces presented in Figures 14-16. Finally, considering KIT, there are significant differences in F1 and F2 between female Mount Roskill speakers and male Mount Roskill and Titirangi speakers. Mount Roskill men have a raised and fronted KIT compared to the Mount Roskill women. While the Titirangi men produce a KIT vowel which is backed and raised compared to the Mount Roskill women. These differences, particularly the height difference between male and female Mount Roskill speakers, are visible in the vowels spaces presented in Figure 15.

6.1.2.2 Age Effects

This section presents the results comparing younger and older female speakers. Given the only gender effect was for GOOSE– which is unreliable in this dataset - using only female speakers to consider age effects is appropriate. There were significant differences for THOUGHT, TRAP, DRESS, and KIT as seen in Table 10 below.

Table 10.
Null model (Young Women) vs. (type-1)*AgeO

	Degrees of Freedom	AIC Differences	Log Likelihood Ratio	p-value
THOUGHT	10	593.8091	16.39571	0.0003
TRAP	10	682.2443	30.3482	<.0001
DRESS	10	811.3826	25.18429	<.0001
KIT	10	677.1308	10.26638	0.0059

Note. AgeO= Older Titirangi speakers

As seen in figures 13-16 in section 6.1.1, TRAP and DRESS are significantly lowered and retracted for all young speakers. KIT is also lowered and retracted for the younger group and THOUGHT is lowered and fronted. This is apparent in the young female vowel plots from Figures 14-16 *(right)*, where THOUGHT is lower than FLEECE for young speakers, but the about the same height for older speakers. These results are presented in Table 11

Table 11.
Null model (Young Women) vs. (type-1)*AgeO – t-tests.

		Value	Std.Error	DF	t-value	p-value
TRAP	typeF1 ‡	6.12185	0.07083	506	86.42825	<.0001
	typeF2 ‡	11.62142	0.05591	506	207.86644	0
	AgeO ‡	-0.78419	0.13064	23	-6.00257	<.0001
	typeF2:AgeO ‡	1.31596	0.16891	506	7.79075	3.806e-14
DRESS	typeF1 ‡	4.78725	0.07143	580	67.02009	<.0001
	typeF2 ‡	12.50552	0.06255	580	199.91875	0
	AgeO ‡	-0.85677	0.13161	23	-6.50981	<.0001
	typeF2:AgeO ‡	1.09710	0.18391	580	5.96548	<.0001
KIT	typeF1 ‡	4.94379	0.06031	304	81.96689	<.0001
	typeF2 ‡	11.02247	0.09642	304	114.31389	<.0001
	AgeO*	-0.30993	0.11044	23	-2.80628	0.01
	typeF2:AgeO	-0.02793	0.21858	304	-0.12776	0.8984238
THOUGHT	typeF1 ‡	3.96960	0.05105	370	77.75456	<.0001
	typeF2 ‡	7.29940	0.09499	370	76.84456	<.0001
	AgeO ‡	-0.44123	0.09415	23	-4.68618	0.0001018
	typeF2:AgeO*	-0.26980	0.12901	370	-2.09135	0.0371782

6.1.3 Monophthong Plots: Conversational Speech

This section presents the monophthong data for casual speech for two subgroups within the Auckland Voices database. These suburbs were selected for reasons given in Section 5.2.2. In addition looking at conversational speech provides some data on vowels that have low token numbers in read speech, NURSE, GOOSE, and FOOT. Token numbers for conversational speech are given in Appendix D (Table D8).

6.1.3.1 Mount Roskill Young Women

As a result of the different types of data, the continuous speech data cannot be compared with read speech. The increased token numbers for most vowels in the casual speech allows for a narrower inspection of the data, however. It allows investigation into inter and intra-speaker variation in this diverse group. Figure 17 below shows conversational monophthong data from the Mount Roskill women plotted on an F1/F2 plane.
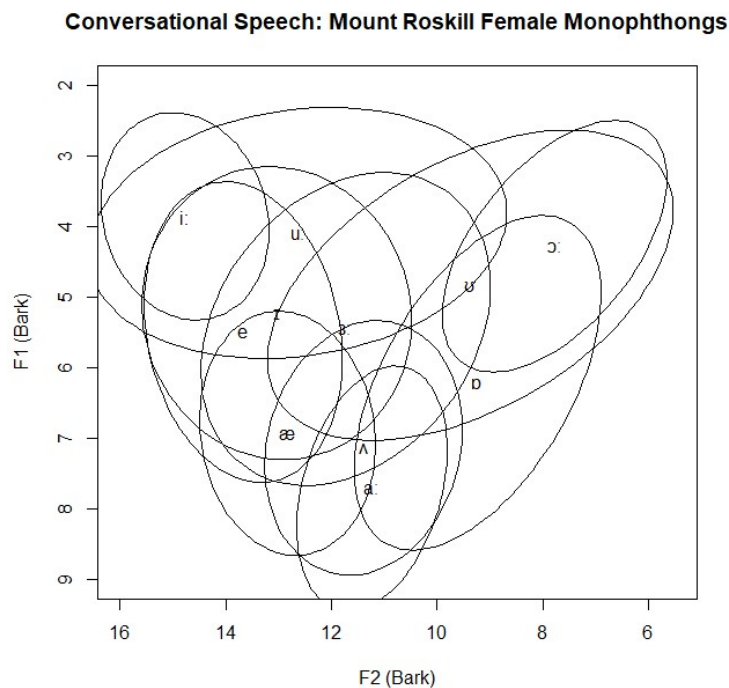

Figure 17. Ellipse plot in F1/F2 space showing vowel centroids of young female speakers from Mount Roskill. Data is from conversational speech. The boundary of each ellipse indicates a 95% confidence interval.

Overall the centroids of these speakers are fairly consistent with read speech findings. TRAP and DRESS are lowered, as is NURSE. KIT is lowered, but is somewhat further forward than might be expected. No explanation is offered for this as there is not enough data to compare; it is possible, however, that this is the result of alternate pronunciations of this vowel emerging during conversation speech which do not appear during conservative read speech. START and STRUT form a long short pair as expected in New Zealand English, THOUGHT is raised and LOT is somewhat unremarkable; although some speakers have large height (F1) variation which is further discussed below. It appears GOOSE is fronted towards FLEECE, but this is likely to be impacted by the frequency of /j/ before GOOSE in English which causes a substantial increase to F2 values. Sociolinguistic factors predict that Mount Roskill women should display signs of extreme inter and intra speaker variation, particularly in conversational speech. In Figure 17 the ellipses for LOT, DRESS and FOOT suggest huge variations in production of these tokens. LOT varies in height, as does DRESS suggesting variations of each present in the data. DRESS variation is not unexpected given the movement of front vowel discovered in section 6.1.1. The variation of LOT in conversation speech, but not read speech could indicate that the New Zealand English pronunciation is considered the conservative or "correct" form, but without conversational data from all suburbs, not too much emphasis is placed on that claim. Rather it is simply evidence of "extreme variation" within this group. FOOT centralization is complicated and cannot be accurately analyzed without more data. FOOT has a notoriously short duration in NZE (see Section 2.2.3) and hence is heavily impacted by its environment. It is possible that the huge variation seen in Figure 17 is the result of centralization in labial and consonantal environments, and retraction in velar consonant environments. An analysis of this vowel in neutral contexts would be required to explore the position of FOOT amongst Auckland speakers.

Given the recent diversification of Mount Roskill, it was expected that unusual combinations of dialectal features might emerge amongst speakers from this suburb. This is already visible in the combination of a New Zealand English vowel space with the obviously non-New Zealand feature of rhoticity performed by two speakers. To explore diversity amongst this group, individual speakers need to be considered. This section presents monophthongs of individual

70

speakers from Mount Roskill conversational speech, plotted in the F1/F2 space. To save space only some speakers of interest are considered here, the remaining speakers can be found in Appendix F *(Figures F1-6)*.
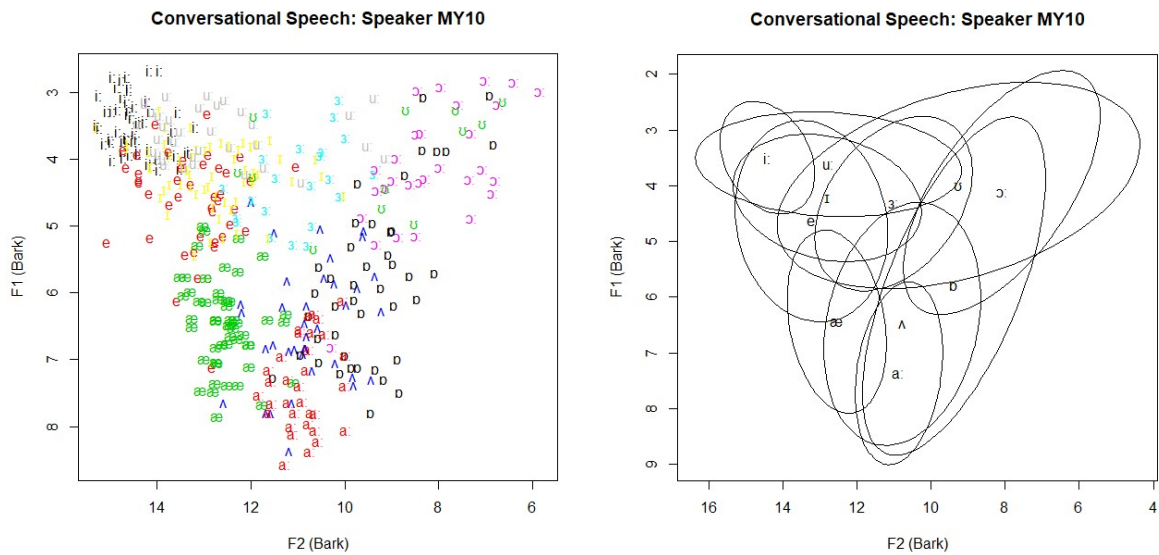
Figure 18. Ellipse plot showing vowel tokens (*left*) and centroids (*right*) of female speaker MY10 from Mount Roskill. The boundary of each ellipse (*right*) indicates a 95% confidence interval.

Figure 18 above shows individual vowel tokens (*left*) and vowel centroids (*right*) for speaker MY03. This speaker has a large F1 spread for LOT, as well as many centralized THOUGHT tokens alongside extremely retracted tokens. Speaker also produced many tokens of STRUT which approach schwa, and tokens of DRESS which are both lowered, and raised.
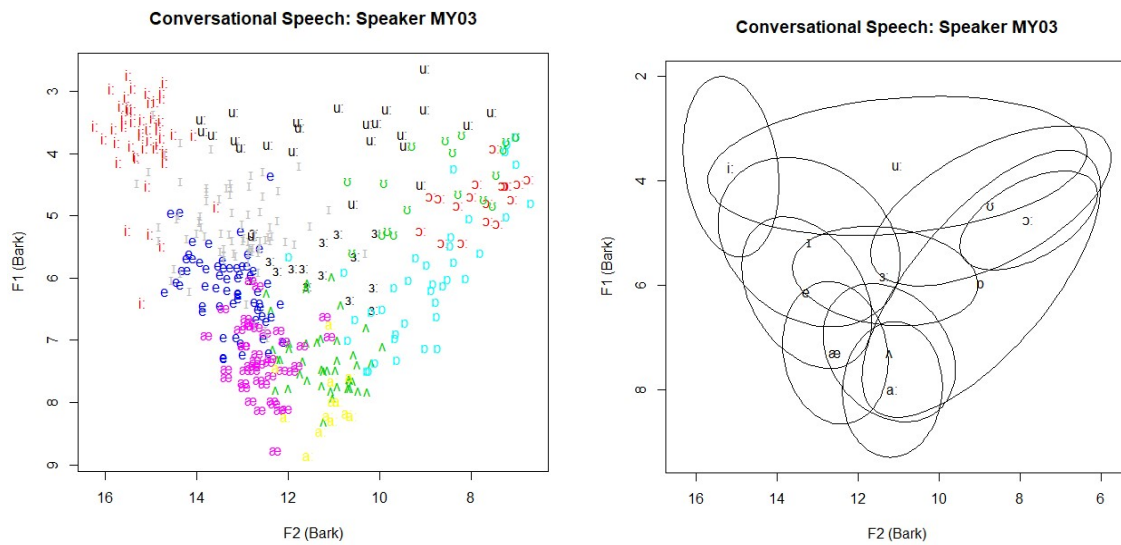
Figure 19. Ellipse plot showing individual vowel tokens (*left*) and centroids (*right*) of female speaker MY03 from Mount Roskill. The boundary of each ellipse (*right*) indicates a 95% confidence interval.

Figure 19 above shows individual vowel tokens (*left*) and vowel centroids (*right*) for speaker MY03. This speaker has a large F1 spread for LOT, as well as a wide range of different KIT realizations from raised and fronted, to retracted and lowered. Speaker also produced some particularly low tokens of FLEECE.

The plots in Figures 18 and 19 show that intraspeaker variation is present in Mount Roskill female speech. This is in line with the vowel plot presented in Figure 17. Without further data, it is not possible to determine whether intra-speaker variation is more prominent amongst Mount Roskill women. Impressionistic evidence, however, suggests that, at least for phonetic features, this is likely to be the case.

6.1.3.1 Papatoetoe Young Men

This section presents monophthongs from conversational speech produced by young men from Papatoetoe. These speakers were selected for analysis for reasons given in Section 5.2.2. As a result of the different types of data, the continuous speech data cannot be compared with read speech. The increased token numbers for most vowels in the casual speech allows for a narrower inspection of the data, however. First an overall description of the data is given, and then individual speakers are considered to outline the process of front vowel lowering in Auckland.
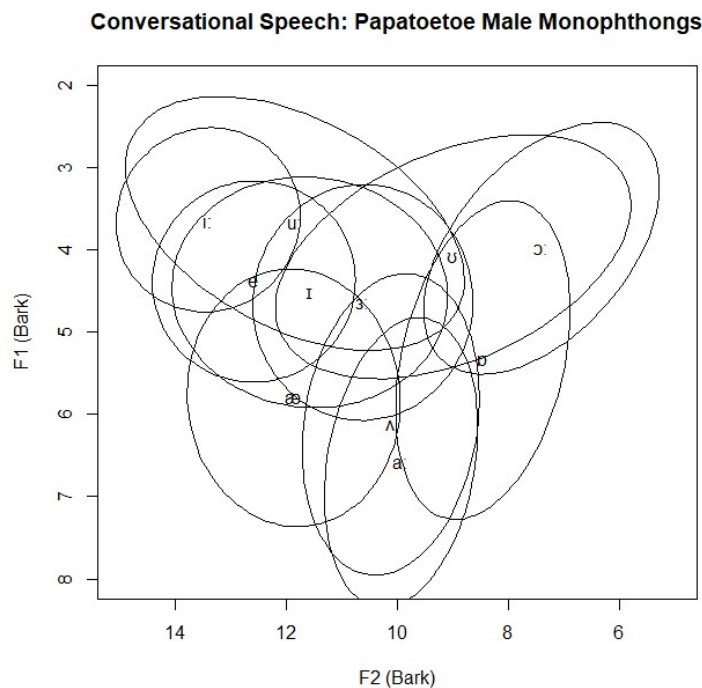
Figure 20. Ellipse plot of F1/F2 space showing vowel centroids of young male speakers from Papatoetoe. Data is from conversational speech. The boundary of each ellipse indicates a 95% confidence interval.

The centroids presented in Figure 20 above, largely confirm the findings from read speech in section 6.1.1. There is some evidence that DRESS is more raised than in read speech, but as the two data types cannot be compared it is not possible to confirm this. Nevertheless, DRESS is

lowered, as are TRAP and NURSE. GOOSE is fronted, although as with the Mount Roskill women above, this is likely environmentally influenced. FOOT is highly variable, supporting the analysis above that data from neutral contexts is required to consider this vowel. LOT varies significantly in F1, as do START and STRUT. A large amount of the variation seen is likely due to anatomical differences between speakers, as with read Papatoetoe speech. One speaker in particular has a compact vowel space which impacts F1 variation for low vowels in particular. KIT is lowered but somewhat fronted, it is possible this is "Pasifika" influence, but without conversational data from all suburbs it is not possible to confirm this.

Figure 21 below considers two individual speakers from Papatoetoe. These speakers display varying levels of participation in front vowel lowering, and hence give an indication of how this change might function. They also show that vowel lowering is a sound change in progress – this is not apparent from read speech data considering the minimal suburb differences in read speech. Only speakers of interest are shown here, remaining individual speaker plots can be found in Appendix F *(Figures F7-13)*.
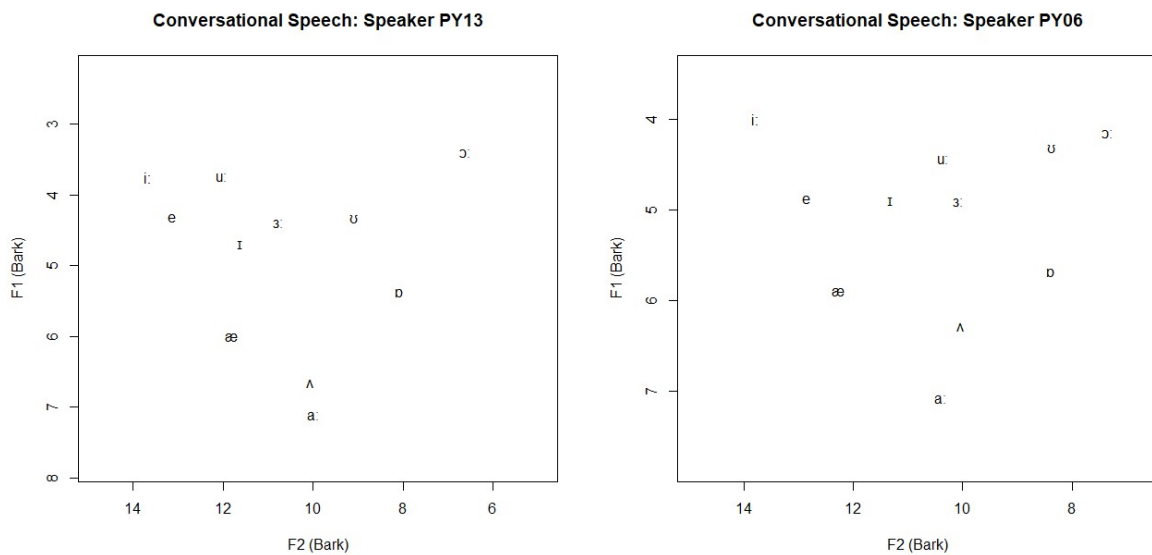
Figure 21. F1/F2 space showing vowel centroids of two young male speakers from Papatoetoe, PY13 (left) and PY06 (right).

Figure 21 shows that while for some individual speakers both TRAP and DRESS are lowered, for others only TRAP is lowered. This suggests the lowering of front vowel seen amongst Auckland Voices speakers began with TRAP. It also indicates the front vowel lowering discovered in this data set is still in progress. This is important given the statistical analysis suggested few inter suburb differences. It is also interesting to note that when DRESS is lowered, NURSE is also lowered, and when DRESS is raised NURSE is raised. Other speakers in this dataset also exhibit varying levels of involvement in this vowel change, these can be found in Appendix F *(Figures F7-13)*. The common feature across all speakers however, is a lowered TRAP vowel.

6.2 Results of Diphthong Analysis

This section presents plots in the F1/F2 plane showing the average trajectory of diphthongs by gender, age, and suburb. Read speech results are presented first, followed by conversational speech. Trajectories were acquired by extracting the formant tracks between the first and second target times. The trajectories were then time normalized, and all trajectories for each phonetic label averaged. The averaged formant trajectories are plotted here on an F1/F2 space in Bark. Trajectories are superimposed over the relevant point vowel ellipses (FLEECE, START, and THOUGHT) for each suburb. CHOICE has been removed from all diphthong plots including conversational speech because of low token count. In some instances NEAR also has low token numbers and in these cases averages should only be considered a guide. For total diphthong count per vowel/suburb see Appendix D (Table D9). This data suggests there is a sound change occurring amongst younger speakers in Auckland. This is most evident for FACE but changes also impact, GOAT and PRICE.

6.2.1 Diphthong Plots: Read Speech
First, from read speech, the older Titirangi Women are considered. As expected, Figure 22 below displays a vowel space very characteristic of New Zealand English. A visual comparison of the older cohort and Watson et al. (1998) conveys this. For space reasons diphthongs from

Watson et al. (1998) have not been repeated here; they can be found in Figure 3, however. For the older Titirangi women NEAR and SQUARE appear merged, although it is not possible to tell whether the first target corresponds to FLEECE or DRESS given the proximity of these monophthongs for this group. The first target of PRICE is also lowered and retracted, which is standard for New Zealand English. Finally, the first target of GOAT is lowered as is common in southern hemisphere dialects. FACE is rarely commented on in the literature, but there are clear similarities between the older cohort presented here and Watson et al. (1998).
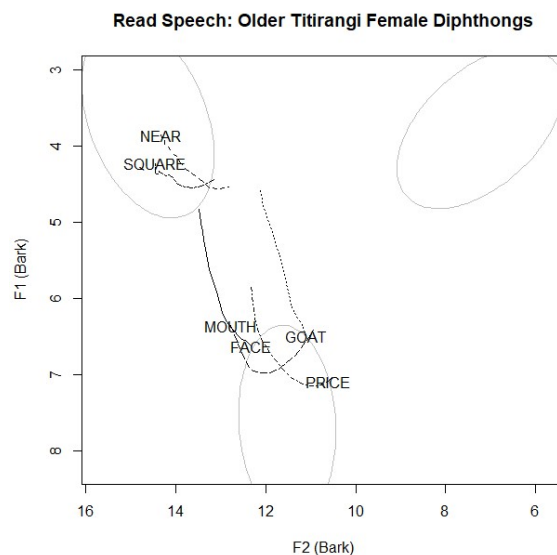
Figure 22. F1/F2 plane showing the average trajectory of diphthongs for older Titirangi female speakers. Formant tracks are extracted between the first and second target, time normalized, then averaged and superimposed over the relevant point vowel ellipses.

The younger cohorts of speakers, however, produce surprising results. As seen in Figures 23-25 below, trajectories for the young Aucklanders differ greatly from the older cohort above. They all, however, demonstrate distancing from this older group in a consistent manner. This suggests that age effects rather than suburb effects are more relevant for diphthong change.

Figure 23. F1/F2 plane showing the average trajectory of diphthongs for Titirangi young male (*left*) and female (*right*) speakers. Formant tracks are extracted between the first and second target, time normalized, then averaged and superimposed over the relevant point vowel ellipses.

Figure 23 above presents averaged diphthong tracks for Titirangi male and female speakers. Male speakers (*left*) in deviate most from the older cohort. The first target of FACE is raised, as is the entire diphthong. The first target of GOAT is also raised, and PRICE is fronted. Female speakers' diphthongs are similar to male speakers', but FACE and GOAT raising is less extreme. The only stable vowels in the diphthong space presented here are NEAR/SQUARE and MOUTH. Of course CHOICE is not considered because of low token numbers.

Figure 24 below presents averaged diphthong tracks for Papatoetoe male and female speakers. These show that male speakers (*left*) deviate from traditional New Zealand diphthongs much like the Titirangi speakers above. Most notably FACE is considerably raised and sits in the space usually occupied by NEAR and SQUARE, although it rises instead of falling. The first target of FACE also notably raised, suggesting a significantly less drastic trajectory than traditional New Zealand English FACE seen in Figure 22. Accompanying this is a raised first target for GOAT, which

77

also differs from previous studies on New Zealand English. PRICE has also fronted, moving towards the space previously occupied by FACE.

Figure 24. F1/F2 plane showing the average trajectory of diphthongs for Papatoetoe young male (*left*) and female (*right*) speakers. Formant tracks are extracted between the first and second target, time normalized, then averaged and superimposed over the relevant point vowel ellipses.

Not only is the first target of FACE raised, but the entire vowel is closer than its traditional New Zealand English counterpart. The female Papatoetoe speakers (right) have undergone similar changes to men, but these changes are considerably less drastic. The first target of GOAT is not raised, and PRICE is more similar to the older cohort than to the male speakers. FACE raising is also less drastic that the corresponding male diphthong. This is indicative of a possible relationship between gender the diphthong changes for young speakers.

Figure 25 below presents the averaged diphthong tracks of Mount Roskill male and female speakers. The Mount Roskill data suggests male and female speakers show the same general

tendencies. Not only is the first target of FACE raised for both groups, but the entire vowel has risen into the close front area of the vowel space for both male and female speakers. The first target of GOAT is raised, and PRICE is somewhat fronted. NEAR/SQUARE and MOUTH, however, seem stable.
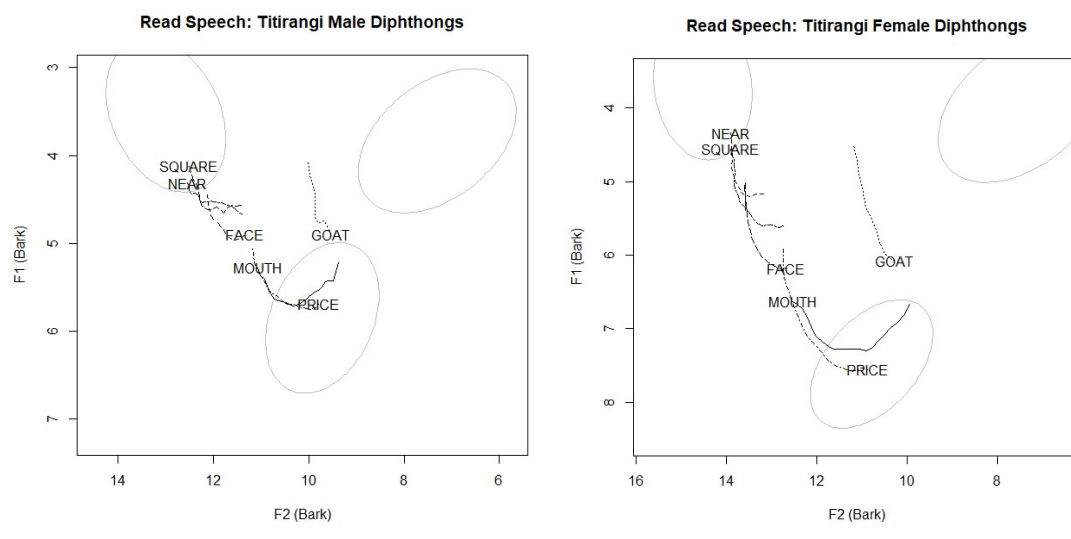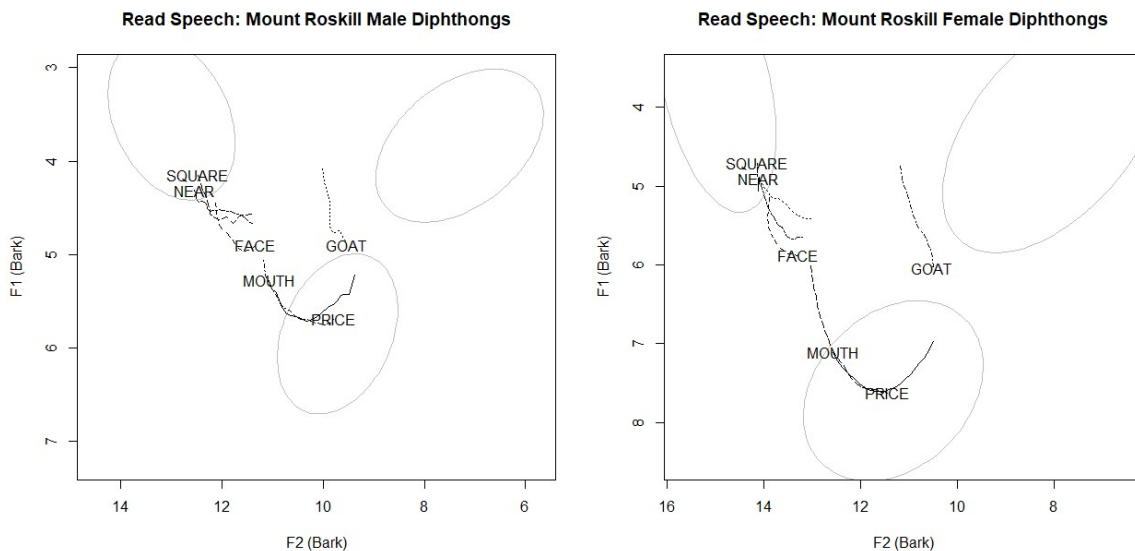

Figure 25. F1/F2 plane showing the average trajectory of diphthongs for Mount Roskill young male (*left*) and female (*right*) speakers. Formant tracks are extracted between the first and second target, time normalized, then averaged and superimposed over the relevant point vowel ellipses.

Given the data above, the most important change is an age effect between older and younger speakers. Compared to the older cohort, younger speakers have a noticeably raised first target for FACE and GOAT vowels. In addition FACE seems to be rising towards the close front area of the vowel space, which is traditionally occupied by NEAR and SQUARE. There is a possible connection between FACE and GOAT, as where one has raised so has the other. In addition, for groups where FACE has risen, PRICE has fronted. This suggests that the diphthongs of young Aucklanders are undergoing some sort of shift relative to traditional NZE which is impacting the entire diphthong system. Cross suburb differences also support this analysis. While diphthongs are

79

shifted for all young groups, changes appear less advanced for Papatoetoe and Titirangi young women, suggesting the change is still in progress, and also indicating some sociolinguistic factors contributing to change. These are further discussed in section 6.2.2. Only NEAR/SQUARE and MOUTH are unaffected by this change, suggesting these are stable features of NZE in Auckland. Although it should be noted there is no evidence of speakers using stigmatized variants of these vowels discussed in section 6.2.2.2. Given that the front monophthongs in Auckland appear to be lowering, it is unusual that diphthongs are rising. It is unclear what is driving this change, but possibilities will be discussed in chapter 7.

6.2.2 Diphthong Plots: Conversational Speech

In this section averaged diphthong tracks from conversational speech data are presented for male Papatoetoe and female Mount Roskill speakers. These secondary results are presented to support read speech data.
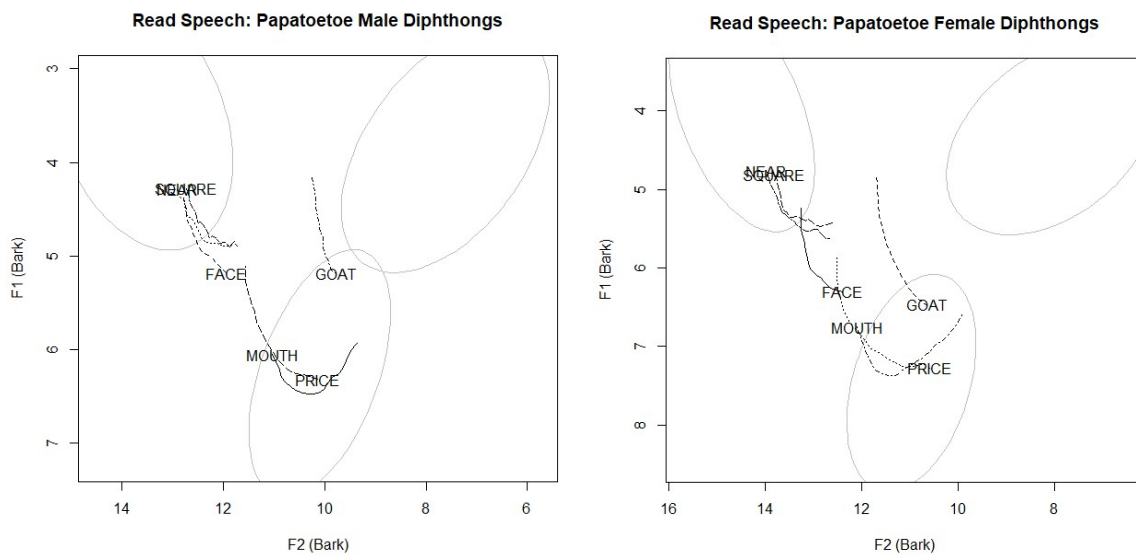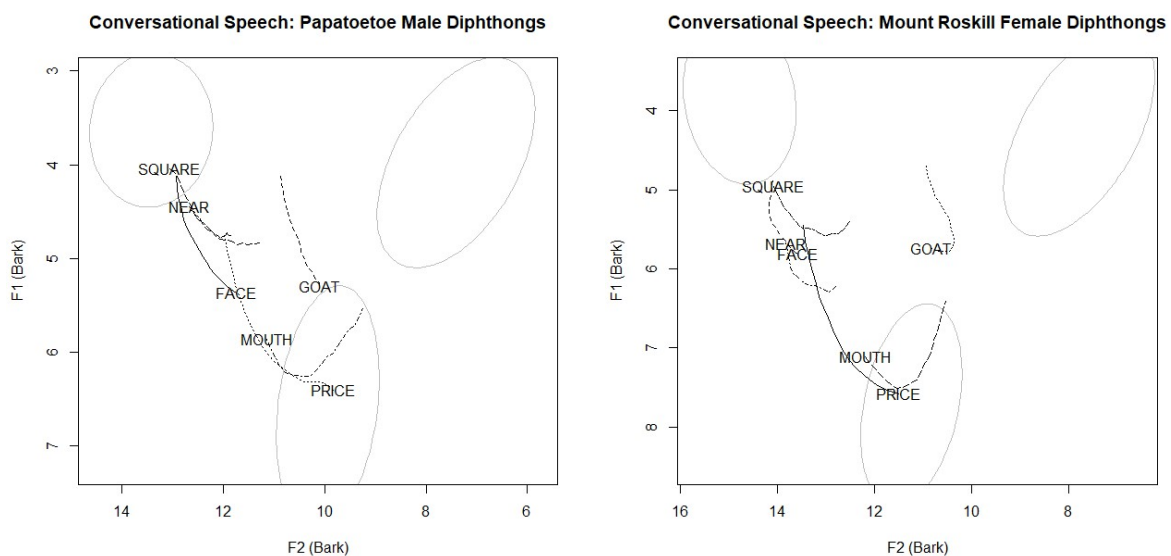

Figure 26. Conversational Speech: F1/F2 plane showing the average trajectory of diphthongs from Papatoetoe male (*left*) and Mount Roskill female (*right*) speakers. Formant tracks are extracted between the first and second target, time normalized, then averaged and superimposed over the relevant point vowel ellipses.

80

First, Mount Roskill Young female speakers above *(right)* largely conform to the patterns identified in read speech above. FACE and GOAT are raised and PRICE is fronted. Interestingly there is separation between NEAR and SQUARE, and it is particularly unusual that NEAR is lower than SQUARE. This could simply be a result of environmental factors, unusual individual speaker variation, or data preparation methods, however. Given token numbers in read speech are insufficient to provide commentary, it will be treated here as an anomaly rather than an actual change. There are some further inconsistencies amongst this group which cannot be explained by low token numbers; these are explored in Section 6.2.2.1 below.

The young male Papatoetoe speakers in *Figure 26* above *(left)* are interesting as diphthong tracks differ fairly dramatically from the same diphthongs in read speech; even though that Mount Roskill speakers (*right*) suggest that the changes identified in the read passage are accurate. For the young male speakers the first targets of both FACE and GOAT are less raised, although the second target of FACE is still raised to align with the first target of NEAR and SQUARE as in read speech; PRICE is still fronted, however. NEAR and SQUARE appear to be merged, with any differences between the two likely to be environmental. It is unusual that in conversational speech Papatoetoe young males produce diphthongs more in line with "New Zealand English" diphthongs seen in Figure 3. Conversational speech is less formal than read speech; and in read speech, speakers are more likely to produce what they consider conservative variants of vowels. Here Papatoetoe speakers, in read speech, are producing diphthongs which are more advanced in proposed changes. This suggests that for these speakers FACE and GOAT raising are considered more conservative variants than the lowered New Zealand counterparts. To confirm this hypothesis conversational data from all other groups would need to be compared, however.

6.2.2.1 Mount Roskill Diphthong Variation

Some of the averaged diphthong trajectories plotted for the Mount Roskill women are not smooth trajectories as one might expect if all vowels were moving in generally the same direction. This can be caused by low token numbers, for instance NEAR in Figure 23 *(left)*. This explanation does not explain rough trajectories such as FACE and GOAT in Figure 26, however. These vowels tend to have high token numbers in our data.

Here this phenomenon is explored by considering diphthong trajectories from conversational speech by individual Mount Roskill women. It seems that some speakers are using multiple variants of diphthongs from their complex linguistic backgrounds. These often move in vastly different trajectories than New Zealand English diphthongs, and because data is labeled phonemically rather than phonetically can result in rough average diphthong trajectories. This is important as it supports the hypothesis that young Mount Roskill women are the most linguistically diverse group of speakers by showing intraspeaker variation.
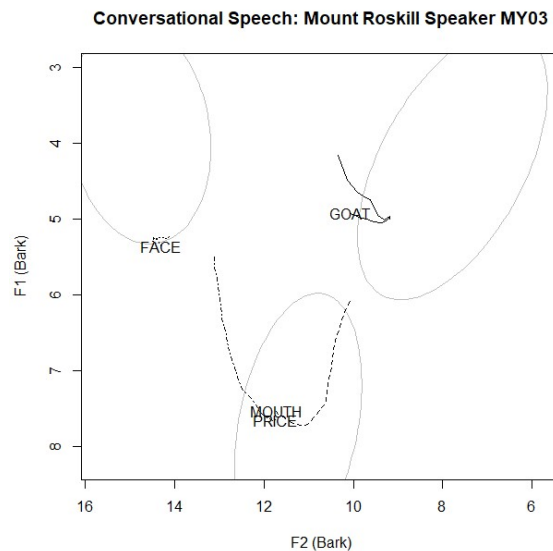
Figure 27. Conversational Speech: F1/F2 plane showing averaged diphthongs tracks from speaker MY03. Diphthongs are plotted on an F1/F2 space, and super imposed over the point vowel monophthong ellipses from each suburb.

82

Figure 27 above shows averaged diphthong tracks of a single Mount Roskill female speaker. The vowel of interest is FACE which appears to double back on itself. Further inspection in Figure 28 below showing individual diphthong tracks plotted as a function of time suggests this speaker is not producing FACE as would be expected. It is significantly more monophthongal, and in some instances falls instead of rising. The diphthong FACE in NZE should have a rising F2 value and falling F1, which is the general pattern for speaker MY01 *(right)* – a speaker from the same cohort. Speaker MY03, however, does not follow this pattern. While a few tokens have a rising F2 and falling F1 most seem to undergo little change in value. This suggests monophthongization. Further inspection also shows that the speaker produces some tokens which in fact fall instead of rising. Since this is difficult to see on the plot in Figure 27, spectral examples are provided in Appendix G.
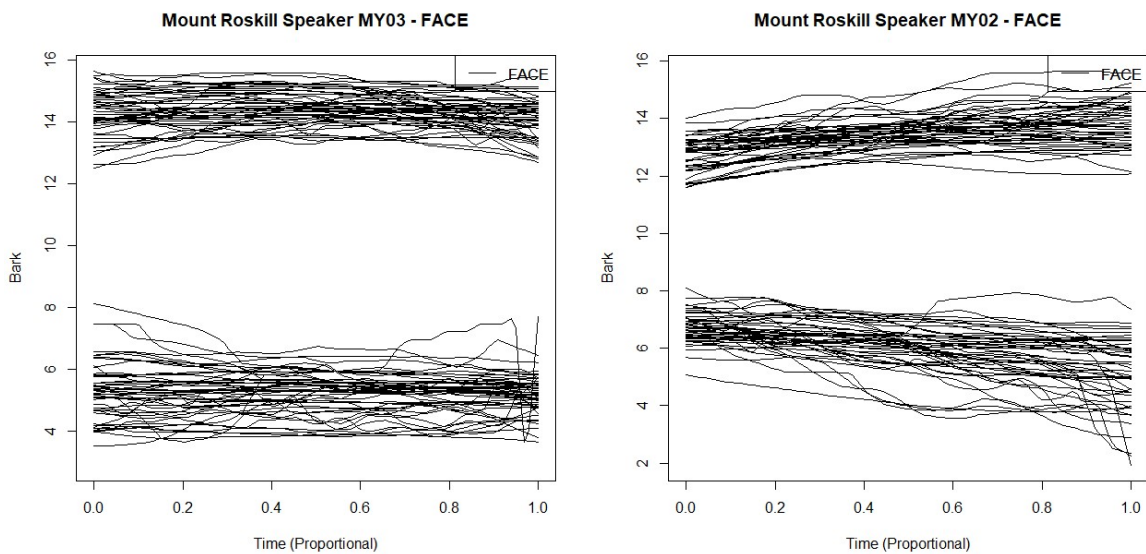
Figure 28. Individual FACE diphthong tracks plotted as a function of time. Young Mount Roskill speakers MY03 (*left*) and MY01 (*right*).

83

This speaker's GOAT vowel is also somewhat monophthongal as seen in Figure 29. It is also often noticeably rounded, more like a THOUGHT vowel than a diphthong.

Mount Roskill Speaker MY03 - GOAT
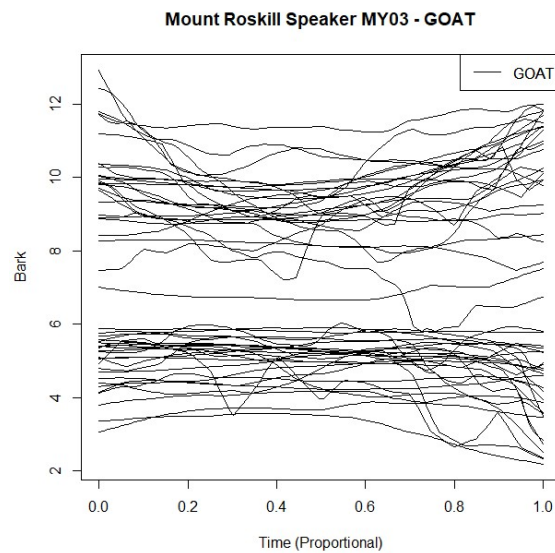
Figure 29. Individual GOAT diphthong tracks plotted as a function of time from female Mount Roskill speaker MY02.

This is not the only unusual variation found for the GOAT vowel amongst this group of speakers; for instance Figure 30 below. Instead of F2 remaining steady as would be expected for this vowel, the trajectories for some tokens produced by speaker MY06 (*left*) rise towards the second target. Speaker MY10 (*right*) produced two variants closer to monophthongs LOT and THOUGHT than to the NZE diphthong, resulting in odd looking formant tracks which do not conform to a general pattern.
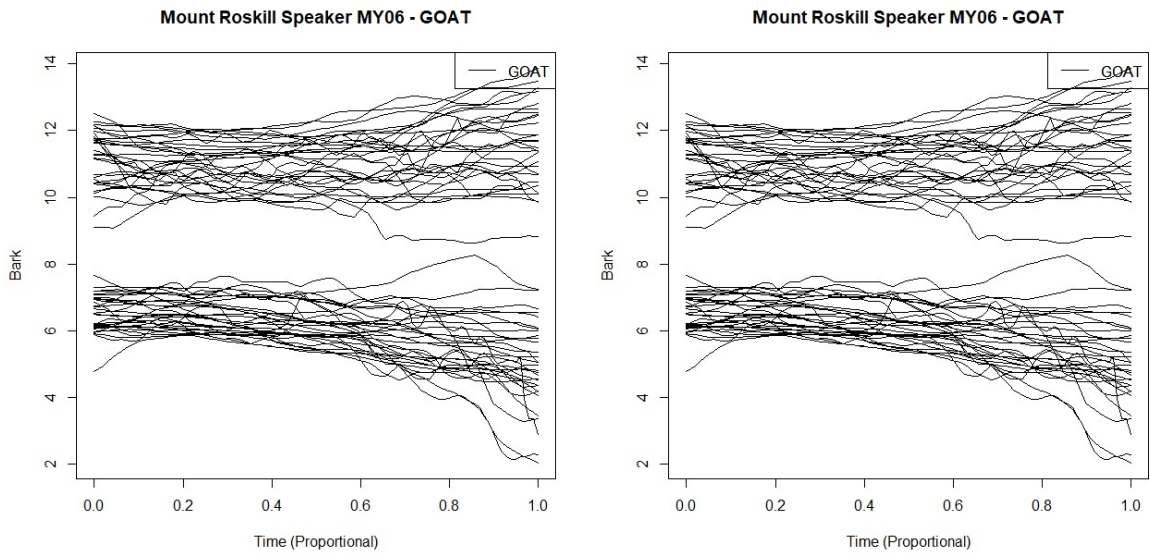
Figure 30. Individual GOAT diphthong tracks plotted as a function of time. Young Mount Roskill speakers MY06 (*left*) and MY10 (*right)*.

These variants likely account for the unusual shape of the diphthong tracts for some vowels in Figures 26 (*right)* and 27. They also provide evidence that speakers in this suburb are using variants of diphthongs that are not directly inherited from New Zealand English. This section does not aim to provide a comprehensive description of all the unusual variants found amongst this diverse group of speakers. For instance one speaker produces MOUTH with a heavily rounded second target. This is simply supporting evidence that there is dialect contact and mixing in Mount Roskill that is impacting the speech of these women. It is unclear whether this extreme variation is driving vowel change in Auckland – but this will be discussed further in Section 7.3.3.

6.3 Vowel Duration

Recent research has considered the impact of vowel length on front vowel movement in New Zealand English (Warren, 2017). Most noticeably, Warren (2017) suggested that the length difference between FLEECE and DRESS was negligible and hence a phonemic distinction based entirely on length was unlikely. In addition the short vowel TRAP is considerably longer in New Zealand English than other short vowel counterparts. In this section the duration of monophthongs and diphthongs in our data set is compared across suburb, gender, and age. Read speech is presented first followed by continuous speech. Duration and standard deviation values are presented in milliseconds (ms). The monophthong FOOT and diphthong CHOICE have been excluded from this analysis because of low token numbers. In addition vowel tokens in syllables which contain a preceding or following approximant (/w, j, l, r/) have been excluded. Phonetic boundaries between sonorants and approximants are unclear due to their similar formant structure (Harrington & Cassidy, 1999). There is even some debate the appropriateness of segmenting approximants and vowels (Harrington & Cassidy, 1999). Given that analyzing vowel duration requires accurate phoneme boundaries, it is best to exclude vowel tokens near approximants from analyses of duration.

Table 12

Short Vowel Duration in Milliseconds by Vowel and Suburb/Gender

Suburb/ Gender	DRESS		TRAP		KIT		STRUT		LOT	
	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev
TOF	105.45	41.87	123.27	55.54	80.97	22.80	90.93	22.31	96.07	33.65
TYF	103.87	30.40	103.31	36.96	83.18	19.41	91.34	31.18	94.73	27.84
TYM	95.53	35.61	101.53	35.73	74.82	19.19	89.49	29.65	86.02	18.01
PYF	105.16	40.84	108.72	44.30	75.06	20.05	115.58	50.66	90.56	28.17
PYM	87.69	30.08	104.56	44.43	69.22	19.50	97.00	32.37	91.01	27.86
MYF	98.22	33.32	108.14	44.24	80.03	23.81	113.13	46.77	96.87	27.09
MYM	95.90	31.80	119.65	62.29	84.39	31.41	93.62	44.78	89.81	32.63
PYM - C	91.54	33.98	130.91	74.52	81.68	45.99	94.37	43.62	109.48	34.50
MYF - C	80.23	31.78	120.61	51.40	75.72	34.47	94.34	40.67	112.85	51.17

Note. Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female (PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female.

Table 13.

Long Vowel Duration in Milliseconds by Vowel and Suburb/Gender

Suburb/ Gender	FLEECE		START		THOUGHT		NURSE		GOOSE	
	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev
TOF	159.07	51.29	186.79	79.16	153.59	56.66	178.16	67.61	164.89	39.61
TYF	135.88	39.94	172.92	73.61	153.50	44.92	160.54	40.03	118.20	21.30
TYM	134.83	51.01	157.20	43.51	135.76	38.19	157.69	52.49	150.48	51.49
PYF	169.40	63.69	161.35	54.27	149.23	43.97	154.28	41.63	190.59	92.01
PYM	139.98	56.85	166.15	98.85	139.82	42.72	119.78	37.38	169.84	90.45
MYF	154.92	64.24	159.25	61.97	154.76	53.18	159.42	39.58	147.41	60.34
MYM	140.40	48.19	135.98	45.06	132.97	47.52	157.84	37.89	136.11	27.95
PYM - C	130.77	65.17	144.95	55.31	145.67	62.18	105.08	44.48	166.24	85.57
MYF - C	121.36	67.59	148.14	56.11	133.72	74.96	98.02	38.36	151.67	117.53

Note. Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female(PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female, Papatoetoe Younger Male – Conversational (PYM – C) & Mount Roskill Young Female – Conversational (MYF – C)

Tables 12 and 13 above show the average duration of short and long monophthongs respectively. The length of vowels in read speech reveals nothing surprising. The vowels which are traditionally "long" are longer than the vowels which are traditionally considered "short". It is interesting that DRESS is notably shorter than FLEECE for all groups, considering recent claims that differences in length between these two vowels are negligible in New Zealand English (Warren, 2017). TRAP is the longest short vowel, but is not nearly as long as FLEECE. STRUT is also longer than might be expected for two groups, Papatoetoe young women, and Mount Roskill young women, but with a large standard deviation it is unlikely to signify any sort of change. In general it seems as though older speakers produce longer tokens than young speakers, but this is likely to a result of speaking style. Older speakers tended to be calmer and more relaxed whilst reading, and took their time. Young speakers on the other hand tended to be a little more rushed resulting in a faster speech rate.

Conversational data provides some interesting results which are consistent across both groups analyzed. TRAP is the same length as FLEECE for both female Mount Roskill speakers, and male

Papatoetoe speakers. DRESS is particularly short, especially compared to FLEECE. NURSE is also significantly shorter than might be expected. This is interesting given the height correlation that often occurs between the two vowels. All other vowels are unremarkable. Although these results are interesting they do not provide much insight without more conversational data to compare.

Table 14 below presents the duration of 6 NZE diphthongs in milliseconds. CHOICE is excluded because of low token numbers. NEAR and SQUARE are presented separately even though they are likely merged for most speakers. This is simply because prior analysis has treated them separately.

Table 14.
Diphthong Duration in Milliseconds by Vowel and Suburb/Gender

sub/ gen	NEAR		SQUARE		FACE		PRICE		MOUTH		GOAT	
	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev
TOF	137.4	43.3	163.9	48.7	181.6	71.6	184.8	46.9	161.4	38.3	146.6	41.6
TYF	122.6	33.5	170.0	66.8	158.7	48.6	161.4	48.1	182.7	73.1	151.0	36.1
TYM	121.7	40.3	129.6	34.5	162.2	48.7	148.3	39.4	142.4	56.7	128.9	42.9
PYF	134.0	52.7	173.6	50.7	175.7	66.1	164.9	47.9	190.2	50.7	146.9	32.3
PYM	130.7	42.8	205.7	116.5	168.4	68.3	153.6	41.4	191.7	73.2	142.1	45.3
MYF	115.3	39.7	177.4	58.6	190.3	77.4	166.1	54.2	191.6	59.3	148.6	35.5
MYM	122.7	28.8	145.8	43.5	138.5	57.0	131.8	26.7	166.9	55.3	135.2	35.3
PYM-C	101.6	26.1	200.1	98.7	134.3	49.0	162.9	69.8	162.0	71.0	140.1	63.4
MYF-C	140.2	112.4	199.5	75.2	154.0	69.4	166.4	76.7	159.9	63.2	177.0	89.1

Note. Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female(PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female, Papatoetoe Younger Male – Conversational (PYM – C) & Mount Roskill Young Female – Conversational (MYF – C)

Diphthong durations from both read and conversation speech also reveal nothing particularly surprising. For most groups NEAR is notably shorter than SQUARE but this is likely due to vastly different token counts (see Appendix D (Tables D9-10)). Differences in duration seem fairly unpredictable and there are no clear patterns of change. Overall GOAT is the shortest diphthong, but there are no differences between older and young speaker groups. Given figures 14-16

from Section 6.2, shortening of FACE for young speakers might have been expected, but there is no evidence here linking raised FACE with a decrease in duration. A couple groups, namely Mount Roskill and Titirangi men have shorter diphthongs than other groups, but this is probably due to a faster speech rate and is consistent with largely short monophthong length for these groups.

6.4 Vowel Onglide

This section investigates vowel onglide for the New Zealand English monophthongs which previous work suggests are diphthongizing. The monophthongs FLEECE, DRESS & GOOSE are often produced with a considerable onglide in New Zealand English (Maclagan & Hay, 2007). This can be investigated by considering how far through vowel production the vowel target occurs. If a vowel is diphthongized or being produced with a significant onglide then the vowel target will be delayed and fall significantly after the vowel midpoint. Warren (2017) suggests that FLEECE is diphthongizing because of pressure from DRESS raising. Our vowel analysis suggested DRESS lowering, however. If DRESS is lowering, then for suburbs where DRESS is lowered a reduced onglide value for FLEECE might be expected. Given this the older cohort of Titirangi speakers should also have the largest onglide. Like duration above, vowels in an approximant environment have been excluded. This is because coarticulation between vowels and approximants impacts formant values and can impact how far through vowel production the steady state occurs, impacting where targets are labeled. Table 15 below shows the onglide of FLEECE, DRESS and GOOSE as a value between zero and one. The closer to one, the later the target falls, and the more diphthongized the vowel. Onglide was calculated for all monophthongs but for brevity only those of interest are presented here. Complete onglide values are presented in Appendix H. Like duration, conversational and read speech are presented together for brevity but cannot be compared because of differences between data sets.

Table 15.
Vowel Onglide by suburb and gender

Suburb/Gender	FLEECE		DRESS		GOOSE	
	Mean	Sdev	Mean	Sdev	Mean	Sdev
TOF	0.63	0.17	0.49	0.14	0.43	0.09
TYF	0.56	0.15	0.50	0.14	0.39	0.11
TYM	0.62	0.19	0.51	0.12	0.40	0.19
PYF	0.60	0.15	0.50	0.15	0.41	0.06
PYM	0.56	0.15	0.51	0.13	0.31	0.15
MYF	0.53	0.14	0.52	0.13	0.43	0.08
MYM	0.55	0.15	0.52	0.14	0.43	0.18
PYM - C	0.56	0.16	0.49	0.14	0.43	0.14
MYF - C	0.48	0.15	0.43	0.15	0.41	0.18

Note. Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female(PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female, Papatoetoe Younger Male – Conversational (PYM – C) & Mount Roskill Young Female – Conversational (MYF – C).

These results suggest that while FLEECE is the most diphthongized for all groups, this diphthongization is reducing amongst young speakers. Furthermore, the older cohort does have the longest onglide, although differences are not substantial. While there were no noteworthy differences between DRESS height (F1) amongst young speakers, there seems to be some difference in FLEECE onglide across suburbs. Mount Roskill speakers and Titirangi young women have the shortest onglide; whereas the Papatoetoe young women and Titirangi young men have the largest. Data from conversational speech suggests that young Mount Roskill women lack an onglide for FLEECE altogether. It should be noted, however, that differences between groups are not particularly large. No cohort has enough of an onglide to be considered a diphthong – and all onglide values are arguably smaller than previous studies on New Zealand English have suggested. These values also show that there is no onglide for DRESS or GOOSE. In fact GOOSE targets all fall before the vowel midpoint. This is unusual, but could simply be the result of data preparation methods, such as target labeling practices. GOOSE targets are labeled where F2 and F3 are closest or at max amplitude, whereas DRESS and FLEECE targets are labeled at the peak of F2. F3 in this dataset is particularly unreliable and could impact onglide values. In addition GOOSE token numbers in read speech are particularly low.

Chapter VII
Discussion of Results

This chapter considers whether the results presented above provide evidence for a unique variety of "Auckland English" which differs to traditional New Zealand English. First, stable features of New Zealand English in Auckland are discussed (Section 7.1), followed by the two major processes of sound change underway in Auckland – front vowel lowering and diphthong shift (Section 7.2). Next, possible gender, suburb and age effects are considered, given this thesis bases much of its methodological design around sociolinguistic theory that suggests some groups within a community are more involved in sound change than others (Section 7.3). The following Section (7.4) combines these analyses; ultimately culminating in a discussion about whether English in Auckland is the same or different to regional New Zealand English.

7.1 Stable features of New Zealand English in Auckland

Even if we are considering the existence of an Auckland variety of New Zealand English, it is still important to illustrate that it is typologically a variant of New Zealand English. The phoneme inventory remains the same, and there is evidence that many features of New Zealand English are stable. In the monophthong system START and STRUT retain their long/short distinction but are otherwise identical. Centroids for LOT also do not vary. KIT remains retracted and lowered for all speakers as expected in New Zealand English. While NURSE may be lowering with DRESS it retains its characteristic rounded quality. THOUGHT has lowered for younger speakers, but not so far as to occupy a different cardinal vowel position. In the diphthong system speakers continue to merge NEAR and SQUARE, and there is no evidence that MOUTH is undergoing any change. CHOICE may also be stable, but low tokens in both read and conversational speech mean that it cannot be accurately analysed in this study. Really it is only the front vowel system, and diphthongs

FACE, GOAT and PRICE which are changing amongst Auckland speakers; if there is an Auckland variety of NZE it is characterised by only these changes from regional NZE.

7.2 Major Vowel Changes

Here we discuss the two major vowel changes identified amongst young Auckland speakers. First front vowel shifts are considered, followed by diphthongs.

7.2.1 Front Vowel Lowering

Much research on New Zealand English has focused on the push chain movement beginning with TRAP that resulted in DRESS raising, KIT retracting and FLEECE diphthongizing (Bauer, 1979; 1992; Trudgill, Gordon & Lewis, 1998; Maclagan and Hay, 2007; Warren, 2017). The Auckland data presented here suggests reversal of this change has been identified among young speakers, however. TRAP and DRESS are both lowering, and FLEECE is becoming less diphthongized. This is unusual given that research suggests sound change reversals don't happen (Bybee, 2003). This section considers the processes involved in this sound change, and why this "reversal" might be occurring.

Front vowel lowering appears to be a sound change in progress, given that data from individual Papatoetoe male speakers presented in Section 6.1.3.2 shows varying levels of participation in this change. Some speakers such as PY13 in Figure 21 have lowered TRAP but raised DRESS; others have lowered realizations of both. If some speakers have only a lowered TRAP, but none have exclusively lowered DRESS, this is tentative evidence that the lowering of front vowels amongst the Auckland Voices speakers is a drag chain beginning with TRAP. This is interesting because research suggests front vowel raising in NZE also began with TRAP. The process of vowel lowering in Auckland is shown in Figure 31 below, along with the possible future movement of KIT.
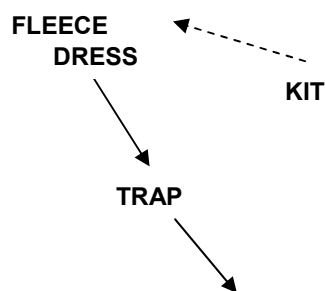
Figure 31. Diagram of front vowel lowering.

In the Auckland Voices data, only TRAP, DRESS and to a lesser extent FLEECE are undergoing sound change. This raises two questions: First, front vowel movement in NZE initially led to the retraction and lowering of KIT. If TRAP and DRESS lowering among these speakers represents a true "reversal" of the NZE front vowel push chain, will it ultimately result in KIT raising and fronting towards FLEECE? Or is centralized KIT an identity marker of New Zealand English that is unlikely to be influenced by front vowel lowering? In the read speech data presented here there is little evidence of KIT raising. Statistically in read speech the young female speakers actually realize KIT lower than the older female cohort. It is also interesting, that Mount Roskill women, the predicted sound change innovators in Auckland, have lowered KIT compared to two male groups. This might suggest that centralized KIT is a stable feature of New Zealand English in Auckland. There is, however, some evidence from the small selection of conversation speech analysed that some Mount Roskill female and Papatoetoe male speakers are fronting KIT. It is unclear, without comprehensive conversational data from all groups, whether this represents the beginning stages of KIT rising and fronting.

Secondly, if TRAP continues to lower and retract, will lowering ultimately impact low central vowels START and STRUT? These vowels were part of the early stages of NZE front vowel raising. As TRAP became closer and more fronted, START and STRUT fronted, but have remained stable since (Gordon, Campbell, Hay, Maclagan, Sudbury, & Trudgill, 2004). Although individual speaker plots are not presented here due to low token numbers, observational evidence suggests that for some speakers TRAP already has a higher F1 value than STRUT. This can be seen in conversational data for Mount Roskill speakers MY03 & MY04 presented in Appendix F

93

(*Figures F2 & 3*). These speakers have a vowel space, at least for short vowels, which is quadrilateral rather than triangular.[3]

Aside from the phonetic process, it is important to consider why is vowel lowering is occurring among these Auckland speakers. Here three possibilities are considered: The first questions whether front vowel lowering in this data marks the beginning of front-vowel lowering New Zealand-wide. The second considers the possibility that this vowel change is the direct result of diversity in Auckland. If this is the case then these changes might indicate the development of a unique "Auckland English". Finally, parallel changes in Sydney Australia are considered given the close relationship between New Zealand English and Australian English.

First, does front vowel lowering in Auckland indicate the beginning of front vowel lowering New Zealand wide? This seems unlikely given recent studies have suggested traditional NZE raising processes are continuing rather than reversing, or even stabilizing. As previously discussed, Warren (2017) found that DRESS and FLEECE essentially overlap in vowel space for both younger and older New Zealand English speakers from Wellington and Hamilton. In this study younger speakers actually had closer realizations of DRESS than older speakers, and more diphthongized FLEECE vowels. This is the reverse of our findings from Auckland Voices speakers. Furthermore, front vowel lowering is present for all young speakers both male and female from all three suburbs in our database. This suggests vowel lowering in Auckland is not a particularly recent phenomenon. Given the proximity of a city like Hamilton to Auckland, if this was a change spreading nationwide, rather than a regionally restricted change I would expect to see evidence of front vowel lowering in Warren (2017)'s analysis.

The second possibility considers whether diversity initiated front vowel lowering in Auckland. One motivation for the acoustic analysis presented here is to investigate how New Zealand English behaves in New Zealand's largest and most linguistically diverse city. It has been previously noted by Maclagan & Hay (2007), that raising of front vowels in New Zealand English,

---

[3] There also is some evidence from Papatoetoe and Mount Roskill conversational speech that TRAP is lengthening. It would be interesting to consider whether continued TRAP lowering and retraction interacts with long START or short STRUT.

particularly DRESS, often goes unnoticed by New Zealand English speakers. This phenomena does however, often cause confusion for speakers of other English dialects who struggle to distinguish between NZE DRESS and FLEECE in pairs such as "ten" and "teen". Hence, in a city like Auckland, which is extremely diverse, it is possible that the need to communicate is influencing vowel lowering for New Zealand English speakers. Notably, DRESS and TRAP lowering among our speakers is resulting in forms which are comparable to more widespread varieties of English such as General American English and RP. If front vowel lowering is caused by linguistic diversity and the influence of northern hemisphere dialects then front vowel lowering should be more advanced in more diverse suburbs, however. In this study, there are only slight differences between suburbs. Possible reasons for the lack of suburb effects are discussed in detail in Section 7.3.2 and 7.4. Finally, if vowel lowering is driven by the necessity to distinguish between DRESS and FLEECE, then the lowering of front vowels did not begin with TRAP as suggested earlier in this section.

It is also interesting that the TRAP and DRESS lowering parallels recent findings from Sydney. New Zealand English is not the only southern hemisphere dialect to have participated in front vowel raising. This sound change has also occurred in Australian English, although with slightly different results as discussed in Section 2.2.1.1. Recent studies from Sydney, however, have discovered TRAP and DRESS lowering and retracting in the vowel space (Cox & Palethorpe, 2007; Cox & Palethorpe, 2008; Elvin, Williams & Escudero, 2016). It seems unlikely that these changes in Auckland and Sydney are related via language contact. Auckland does not have any particular affinity to Sydney that would result in the transference of language features. What they do have in common, however, is extreme diversity. Where Auckland is New Zealand's most diverse city, Sydney is Australia's most diverse city (Australian Bureau of Statistics, 2014). This might support the suggestion that changes in Auckland are caused by language contact with northern hemisphere dialects in a diverse environment.

Finally, it must be considered whether vowel lowering in Auckland indicates that Aucklanders wish to forge their own identities outside of simply being New Zealanders. This could result in

the stigmatization and hence rejection of some New Zealand English forms, such as raised front vowels. This is discussed in more detail in section 7.4, given it is relevant for all vowel changes.

7.2.2 Diphthong Shift

Young Auckland speakers in our database seem to be rejecting some typical New Zealand English diphthong realizations for new variants. The most noticeable change is the raising of FACE. For all young speakers FACE is raised to some extent and its trajectory is shortened. This accompanies PRICE fronting and raising of the first target of GOAT. Much like the front vowel system, it appears that there is some link between these diphthong changes. Plots from Figures 23-25 show a consistent link between raising of the first target of FACE and GOAT. For instance, in read speech young Papatoetoe women have both a lowered GOAT and FACE vowel, but males have raised both (Figure 24). This relationship between FACE and GOAT seems to true for all groups. Differences between young speaker groups suggest this changing is still in progress. Papatoetoe and Titirangi young women in particular have lower realizations of FACE and GOAT than other groups. The first targets are still raised compared the older cohort of speakers, however.

Given most in depth studies of diphthongs in NZE have focussed on NEAR and SQUARE, it is difficult to establish the cause of diphthong shift among these young Auckland speakers. These NZE diphthongs tend to appear stable, with the exception of broad New Zealand English variants discussed in Section 2.1.1. No speakers in the Auckland voices database produce broad NZE variations of any diphthong, however. This could be because of the formal interview setting, but also is likely a result of the urban setting of this study. Broad New Zealand English tends to be associated with rural New Zealanders and therefore is likely to be highly stigmatized in metropolitan Auckland.

If these diphthong changes are indeed part of a system, what might the process of change be? Evidence from Auckland Voices speakers, suggests diphthong shifting may have started with PRICE. Despite the movement of FACE being the most obvious change, it is PRICE which is consistently fronted for all young speakers. For all young speakers PRICE seems to be fronted

towards the traditional NZE FACE vowel, whereas it is retracted for the older cohort. It unclear what may have cause PRICE to front. One possibility is the influence of other English dialects present in Auckland, such as General American English (Wells, 1982a), and Indian English (Gargesh, 2008) which both use fronted variants of the PRICE diphthongs. Perhaps, as with monophthongs, it is the need to be understood in a diverse environment, or even dialect contact and mixing which has started this change. If PRICE movement in Auckland is related to diversity, however, then more advanced fronting/change would be expected in diverse suburbs of Mount Roskill or Papatoetoe.

It is also possible that the data presented here signifies increased stigmatization of NZE diphthongs among young speakers in Auckland. The Papatoetoe male speakers considered in this study, in conversational speech, use diphthongs which are closer to typical NZE realizations than their counterparts in read speech; particularly for FACE and GOAT. If conversational speech is less conservative than read speech, this suggests that our Auckland speakers – at least in Papatoetoe – consider the innovative diphthongs produced by all young speakers in read speech conservative. The stigmatization of NZE diphthongs could be caused by two factors: the first pertains to the Identity of Aucklanders. It is possible that Auckland speakers are rejecting New Zealand English forms in order to forge their own identities as Aucklanders. This is discussed further in the following section. The second reason for stigmatization of New Zealand English diphthongs could simply be a result of communicative needs as discussed above. Like front vowel movement, diphthong changes among the young Auckland speakers are not necessarily towards any new or innovative forms in when considering global Englishes. Both FACE and GOAT are becoming more similar to vowels found in northern hemisphere dialects. FACE for instance is realized as [eɪ] in both American and British dialects (Wells, 1982a), and as the long monophthong [e:] in Indian English (Gargesh, 2008). While these are not identical to the variety found amongst our speakers, they are all more raised than the typical NZE variant. GOAT in RP also has a first target which is raised compared to the NZE diphthong. This is in addition to data from Mount Roskill young women suggesting speakers in this diverse suburb are retaining

diphthong variants from input dialects other than NZE. More data from conversational speech is required to investigate this phenomenon, however.

7.3 Age, Suburb, and Gender Effects

The following sections consider whether there are age, gender and suburb effects present among the Auckland Voices speakers.

7.3.1 Age Effects

Are there age effects between speakers under 25 and speakers over 40? Yes. In fact the most significant differences uncovered by the acoustic analysis are age effects. The younger cohorts of Auckland Speakers have TRAP and DRESS vowels which are significantly lowered compared to the older cohort. THOUGHT and KIT also differ between younger and older groups. KIT is lowered, which is somewhat unusual. Given the systematic reversal of front vowel changes seen here, raising may have been expected. Most importantly, however, despite statistical differences, the young speakers seem to be retaining the iconic centralized KIT vowel that characterizes New Zealand English. THOUGHT is significantly lowered and fronted for young speakers. Given that raised and retracted THOUGHT is a typical feature of Southern Hemisphere Englishes, as discussed in Section 2.2.3, this could also be further indication that young speakers in Auckland are rejecting some typical New Zealand English forms.

Young speaker's diphthongs also differ from the older cohort. Comparing figures 3 and 22 showed that again, like the monophthong data, older speaker's vowel spaces are similar to those presented in previous New Zealand English studies. Despite the aforementioned differences between younger and older speakers, it is important to note that there are many vowels which remain stable across age groups. In the monophthong system this is LOT, START, STRUT and FLEECE[4]. In the diphthong system all cohorts merge NEAR and SQUARE – with plotted

---

[4] FLEECE is more diphthongized for older speakers, but occupies the same space on an F1/F2 plane for both older and younger speakers.

differences likely environment based. As suggested by Watson et al. (2000), some speakers use both variants of NEAR and SQUARE but not phonemically. MOUTH also seems stable and conforms to results from previous studies on New Zealand English.

7.3.2 Suburb and Gender effects

The research questions proposed in Section 3.3 asked two questions: Is there a relationship between gender and language change in Auckland? Also, are there significant cross suburb differences in vowel spaces supporting the MLE study by Cheshire et al. (2013), or does variation and dialect contact function differently in Auckland?

Contrary to expectation, there were minimal suburb and gender effects among the Auckland Voices data. This is divergent to previous work on NZE which suggests a relationship between gender and sound change (Maclagan et al., 1999; Maclagan & Hay, 2007; Gordon & Maclagan, 2001). Based on Labov (1994), young women should lead non-stigmatized change in unstable linguistic environments, and young men should be more likely to pick up stigmatized features in stable linguistic environments. The data presented here also suggests that dialect contact and dialect mixing are functioning differently in Auckland than in the suburbs considered for the London based MLE study (Cheshire et al., 2011, 2013). According to work on MLE, speakers from more diverse suburbs should be leading sound change. Combining analyses about diversity and genders suggests that young women from diverse suburbs are most likely to be sound change leaders. In our data, however, there is not sufficient evidence to suggest any speaker group is leading sound change in Auckland.

The statistical analysis shows significant differences for interactions between suburb and gender for THOUGHT, TRAP, FLEECE, KIT and GOOSE; but these results do not suggest that any group is leading any of the changes. If this were the case we would expect to consistently see evidence of the same group being statistically more advanced in proposed changes; this is not the case. Sociolinguistic theory also suggests that young men and women from Titirangi should

99

be more conservative than speakers from more diverse suburbs. There is no evidence to suggest this in the speech analyzed here. Both male and female Titirangi speakers are participating in the lowering of monophthongs DRESS and TRAP. In fact Titirangi young men have a statistically lower TRAP vowel than the Mount Roskill women who were expected to lead sound change. There is some evidence from the diphthongs presented in Section 7.2.2 that Titirangi and Papatoetoe women are participating less than other cohorts in diphthong shifts (Figures 23 & 24). If diphthongs are shifting, however, they are still considerably more advanced in this change than the older cohort of women. In addition Papatoetoe is a diverse suburb, where change is expected; theoretically Papatoetoe speakers should be more advanced than the Titirangi women in any changes. Mount Roskill women, however, are as advanced as male speakers in diphthong changes. This is evidence suggesting slightly different responses to dialect contact processes in recent (Mount Roskill) and established (Papatoetoe) migrant communities. It could also be argued that the use of shifted forms by young female Mount Roskill speakers, where the young Papatoetoe male group use more traditional NZE forms is further evidence for this. This cannot be confirmed without more data from conversational speech, however.

7.3.3 Young Mount Roskill Women

Sociolinguistic factors suggested that young women from Mount Roskill were likely to be the most variable speakers of New Zealand English. Here we consider that while these speakers might be the most variable in our dataset, this does not imply they are driving language change in Auckland.

Dialect formation theory suggests that dialect contact will result in increased intraspeaker and interspeaker variability (Trudgill et al. 2000). Trudgill et al. also note that "extreme variability" often manifests as unusual combinations of features. Of course evidence for dialect mixing in Mount Roskill is likely to be less extreme than the ONZE data considered by Trudgill et al. (1999). Linguistic diversity in Mount Roskill is not isolated from outside influence like the

speakers considered for the ONZE project. In addition, there are significant differences between the dialect mixture which resulted in New Zealand English and the dialect mixture found in Mount Roskill. Where the dialect mixture of early New Zealand was largely competing varieties of British English, the mixture found in Mount Roskill is likely a mixture of many distinct English varieties from across the world. Englishes such as (but not limited to) Indian, American, Singaporean, British, and Pasifika are all present in Mount Roskill. Both groups however, represent the first generation of speakers born in a new linguistically diverse community.

While the vowel centroids plotted in Figure 15 *(right)* do not differ greatly from any other young speaker group, there is evidence for increased variability amongst these speakers. Among this group, unusual combinations of features were identified; for instance the combination of a rhotic dialect with a New Zealand English vowel space. Intraspeaker variation was also identified; for example in Section 6.2.2.1 we showed that some young Mount Roskill women are using multiple different variants of diphthongs such as FACE and GOAT. In addition conversational speech data was presented in Section 6.1.3.1 showing individual speakers produce multiple variants of some monophthongs. This suggests processes of dialect formation predicted by Trudgill et al. (2000) are occurring in Mount Roskill.

Somewhat surprisingly, however, the in-group variation described above does not seem to correlate to greater advancement in either of the major vowel changes proposed here. Young Mount Roskill women are not more advanced in TRAP or DRESS lowering, and equal young men in diphthong changes. Statistically in terms of changes identified, in some instances they even lag behind other groups. For instance, the Titirangi men have a statistically lower TRAP vowel than the Mount Roskill women, and all male groups have a lowered THOUGHT vowel compared to the women. This suggests that while as expected Mount Roskill women are a diverse group of speakers, at least in read speech they are not leading change as might be expected. It would be interesting in future research to consider conversational data from other groups, to determine

whether there is statistical evidence for Mount Roskill women leading monophthong and diphthong changes.

7.4 Variation in Auckland and "Auckland English"

This section explores how sociolinguistic phenomena might function within our Auckland based data sets and whether the data presented here is evidence of "Auckland English". The nature of vowel change in Auckland is somewhat peculiar, given there are large age effects, but somewhat minimal gender and suburb differences. It is unusual given how sound change usually functions sociolinguistically that young speakers from our homogenous – presumably linguistically stable - suburb are participating in sound change, while the older cohort seem to have no involvement at all. Ultimately there are two possibilities: The first is that we are seeing the beginning of widespread changes in New Zealand English in these young speakers. The second is that these results represent the development of an Auckland specific variant of New Zealand English.

The first possibility, that the changes discovered here represent changes to New Zealand as a whole beginning in Auckland, seems unlikely. Recent studies on New Zealand English have found no evidence of any of the changes described here, and in fact tend to show the continuation of existing changes; for instance, Warren (2017) found the continuation of DRESS raising and FLEECE diphthongization among young speakers in Wellington and Hamilton. There is unfortunately little recent work on New Zealand English diphthongs. Given the results presented here it would be interesting to consider the diphthong space of young speakers from outside of Auckland to confirm that these changes are locally bound. This is not, however, within the scope of this project.

The other possibility is that this data presents evidence of an Auckland variety of New Zealand English. Given that there is no evidence of these changes existing outside of Auckland, this

seems likely. There are however, some problems stemming from the analysis of our data that must be addressed before making such a bold claim about "Auckland English".

First, given the lack of suburb differences, but abundantly clear age effects, suggesting the changes presented here are fairly well developed, why has this change not been previously commented on? This is possibly at least partially due to the lack of studies considering Auckland speakers. It is also possible that the vowel changes discovered here don't tend to be noticed, because they are not really changes towards "innovative" forms. Front vowel lowering and diphthong changes in Auckland are changes away from the innovative NZE/Southern hemisphere forms, but they are not moving towards anything particularly new or unfamiliar. American English, British English (Wells, 1982a) and Indian English (Gargesh, 2008) dialects all tend to have lowered TRAP and DRESS vowels than NZE. These are all likely to be present in the dialect mixture impacting New Zealand English in Auckland, if these changes are influenced by diversity.[5] FACE also seems to moving towards a variant more similar to the RP FACE vowel, and fronted PRICE is present in General American English (Wells, 1982a) and Indian English (Gargesh, 2008). Hence, these changes, while different to typical New Zealand forms, are not forms that New Zealanders are unfamiliar with.

Another concern regarding the possibility of Auckland English is why changes are largely ubiquitous across suburbs, but older speakers do not seem to be participating at all? It is of course possible that the lack of variation across suburbs is the result of speakers from more diverse suburbs such as Mount Roskill behaving conservatively in a research environment; masking variation and differences across groups. If this is the case, then it is important to note, this implies all young Auckland speakers consider many non-New Zealand English features to be conservative. This would be an interesting topic for further research. The other possibility is that suburb and gender differences are minimized because changes towards an "Auckland English" variety are so advanced that dialect levelling has already occurred. In terms of dialect formation processes outlined in Trudgill et al. (2000), where the expectation was that our

---

[5] For many migrants to Auckland, English is a second language; most of these speakers learn northern hemisphere varieties of RP or General American English. Auckland also has a large Indian population, who are likely to speak varieties of Indian English rather than varieties based on RP or American English (Gargesh, 2008).

speakers would be displaying signs of extreme variability between and within suburb groups, in actual fact the changes we are seeing are so advanced that speakers are focussing on particular features and a new dialect is emerging. This analysis is somewhat questionable given there is zero evidence of change amongst the older speakers, however. It seems dubious that the emergence of a new dialect would occur in the span of 20 years without any effect on the older cohort of speakers. Given the only older speakers analysed are from Titirangi, it is of course possible, that looking at data from older speakers from Papatoetoe and Titirangi might reveal evidence of older speakers participating in language change.

In addition, as previously mentioned, Auckland has undergone rapid diversification over the past 30 years. This diversification began with changes to the immigration act in 1987 which stopped immigration from focusing on nationality and immigration as the basis for admission (New Zealand Parliament, 2008); instead admission was granted to anyone who met educational, business, professional, age, or asset requirements. This caused rapid diversification, particularly in Auckland where many migrants settled. For instance, according to Statistics NZ (n.d.*b*) the number of Auckland residents who identified as European (Including NZ European) reduced from 76% in 1991 to 64% in 2006. While the number of resident who identified as "Asian" increased from 6-16% in the same period. This change is significant because most of the young cohorts of speakers considered in this study are in their early to mid 20's. This means they were born sometime during the 1990s, during the years immediately following New Zealand's immigration policy change. Diversification of Auckland was not gradual, and hence these young speakers represent the first generation of New Zealand English speakers raised in this super-diverse city; much like the speakers analyzed in the ONZE project represent the first generation of speakers born in New Zealand (Trudgill et al., 2000). This rapid diversification and increased dialect contact and mixing could have caused rapid changes to the English spoken in Auckland, which the young speakers analysed in this database would have been at the centre of. The older cohort of speakers, however, were all born significantly before the diversification of Auckland. Most would have been in their 20s or 30s when diversification of Auckland began, and would have grown up in a time where regional variations of New

Zealand English did not exist. Given this they are likely to lag behind in any changes occurring in Auckland. More data from older speakers and additional conversation data would provide more weight to this analysis.

Other possibilities for the lack of suburb effects involve the mobility of young Aucklanders, meaning that topographical communities have little relevance for the Auckland speakers considered here. While the young speakers considered in the MLE study (Cheshire et al., 2013) were all teens, our speakers tend to be in their early 20s. While our speakers largely went to primary and secondary school in the relevant suburb, many of them are now in tertiary education or in the workplace. This requires them to travel frequently outside the confines of their suburban communities and interact with New Zealand English speakers from across Auckland. This mobility likely increases dialect contact between speakers from different areas. Also given that 39.1 percent of Auckland residents are not New Zealand born (Statistics NZ, 2013b), the mobility of young speakers means that even speakers from fairly homogenous suburbs have contact with English speakers from a wide range of different linguistic backgrounds. It possible that this mobility might contribute to the dialect levelling that is seen amongst young speakers from different suburbs. In particular it is possible that changes to New Zealand English in Auckland could exist as a result of speakers needing to be understood in a diverse environment. Some changes, for instance DRESS lowering, are to do with features of New Zealand English which are not usually "identity markers" but which often cause confusion for non-New Zealand English speakers.

In relation to why Auckland might be developing its own unique variety, relative isolation from "regional" New Zealand English is another factor which should be considered. While Auckland speakers are certainly not as isolated as the ONZE speakers considered by Trudgill et al. (2000), it is probably true that Auckland residents have less contact with other regions of New Zealand. Auckland is so large that Aucklanders have little need to travel outside of city. Anecdotal evidence from the speakers in the Auckland Voices database suggests that if speakers do travel outside of Auckland it tends to be to tourist destinations such as Queenstown, Taupo or the Coromandel, where it is unlikely that they will have much contact with "regional" New Zealand

English. Interacting with other Aucklanders or tourists who speaker other English dialects instead. This could contribute to localized change amongst Auckland speakers.

Finally we consider the possibility that Auckland speakers are *rejecting* some New Zealand English forms in order to forge their own identities as Aucklanders. There is evidence in our data that some New Zealand English variants, particularly diphthongs are stigmatized for young speakers. It is possible that this stigmatization signifies that young Auckland speakers are intentionally moving away from traditional New Zealand English forms as a way of setting themselves apart from other New Zealanders. In order to explore this possibility more data from conversational speech would need to be prepared and analyzed.

Ultimately, this thesis also cannot provide a definitive answer about the processes contributing to language change in Auckland. These processes are less clear cut that those found in London (Cheshire et al., 2011, 2013), and also do not entirely conform to the processes of dialect formation suggested by Trudgill et al. (2000). It is likely that dialect formation and change in Auckland is affected by a combination of sociolinguistic pressures relating to diversity, identity, and mobility. This complex linguistic situation results in the unusual linguistic scenario presented here, where we see age effects and language change, but not suburb or gender effects. This is important because it shows that language change and variation in a diverse metropolis is not always a clear process where diversity equates to variation which then results in language change.

Given the above we think that it is worthwhile considering whether there is an Auckland variety of New Zealand English developing, but without more evidence such a bold claim should not be made. The data presented in this thesis certainly suggests there is an "Auckland English" but further studies of Auckland speakers would make this claim more reliable. The possibility of an "Auckland English", however, is a major discovery for New Zealand English research, because previous research has focused on the lack of regional variation in New Zealand.

7.5 Papatoetoe and Pasifika English

In this section a brief discussion of Papatoetoe speakers and evidence for "Pasifika English" in our data will be provided. Given the demographics of Papatoetoe (45.7% identify as Pacific Peoples), and recent work on "Pasifika English", it was predicted that evidence of influence from Pacific languages might be found among the vowel spaces of these speakers. There is evidence of many of the features mentioned in Pasifika English studies; for instance, lowering and retraction of DRESS and TRAP, and reduced diphthongization of FLEECE and GOOSE. These features however are not unique to Papatoetoe; they are common to speakers from all three suburbs analyzed. This suggests that some of the features identified as "Pasifika English" vowel features, are likely features of "Auckland English". This is not surprising given that studies on Pasifika English tend to take place in Auckland, and there has previously been a lack of acoustic studies involving Auckland speakers.

This raises the question of whether distinct Pasifika English exists in New Zealand. While we did not find any evidence of vowel features amongst Papatoetoe speakers, there was evidence of some consonant features which have been ascribed to "Pasifika English". These include –TH fronting and –DH stopping (Starks, 2006; Gibson & Bell, 2010), as well as the use of a semi rhotic dialect in some contexts (Starks, 2006). This thesis is not concerned with consonant features, however, and these have not been further analysed.

7.6 Methodological Issues and Future Research

Following the discussion of results, some methodological concerns must be considered. This is followed by a discussion of where future research might be useful.

In regards to methodological concerns, although rigorous data preparation was undertaken, much of the data analyzed was from noisy recordings where formant trackers perform poorly. This noise is unavoidable if natural speech is to be collected, but does result in increased formant track errors. This caused some issues for diphthong tracks as formants tracks tended to

shaky. These were often left uncorrected because they were still tracking along the correct formant, and corrections should only be undertaken if necessary. This, however, meant that formant tracks sometimes did not transition smoothly from the first to second target and this was visible when plotting diphthongs.

There are also some issues with plotting vowels at a vowel target. This was a problem for vowels such as GOOSE which often precedes the onglide /j/; for instance in words such as usually, used and you. Following guidelines from Watson et al. (1998), GOOSE targets are marked at the closest point between F2 and F3. /j/ however raises the second formant of GOOSE. Following /j/ the closest point between F2 and F3 often results in a target with a significantly higher F2 value than is expected. This resulted in GOOSE appearing more fronted than would be expected in our, and makes analysis of this vowel unreliable.

It is also a concern that the read speech analyzed in this study has low token numbers for NURSE, GOOSE, FOOT and CHOICE. This meant that an accurate analysis of these vowels could not be performed. Future research on Auckland English might look at these vowels and how they are changing in Auckland; particularly NURSE which seems to be lowering with DRESS for younger speakers. Although there are conversational speech tokens for these vowels, there is no conversational speech from the older group to compare. In addition CHOICE is infrequent in conversational speech, so we cannot provide any analysis on this vowel.

Another limitation of this study is that due to the method of recruiting there is no attempt to balance for group types. Ethnic identity is self-reported in the interviews collected, but at varying levels of detail. This means we can say that our Mount Roskill speakers are representative for instance of the diverse Mount Roskill community and postulate that variability amongst the Mount Roskill women is the result of this – but we cannot make any explicit claims about the impact of individual speakers backgrounds on their speech. This is true also for the Papatoetoe groups. While we can confirm that the speakers considered are largely from a Pasifika background, we cannot say for example: Person A has X feature because they have a Pasifika background.

Finally a major concern is that the analysis provided here is largely based on a read passage of text with supplementary casual data. Analysis of conversational speech for all groups would provide extra analysis power. Read speech is useful because speakers perform the same piece and each provides a similar number of vowel tokens for each vowel; this helps to control many speaker effects. The downside of primarily relying on read speech is that speakers tend to perform conservatively when reading. Conversational speech is much more casual and speakers are less likely to use conservative forms, but data can be skewed by factors such as varying token numbers. Ideally this study would present an analysis of both conversational and read speech.

Furthermore, evidence from the limited selection of conversational speech analysed suggests opportunities for further investigation. For instance the conversational data provided some evidence that TRAP might be lengthening for young speakers in Auckland. For both Papatoetoe and Mount Roskill groups analyzed, TRAP was equal in length to DRESS. It would be interesting to see if TRAP lengthening is present for other young groups within the Auckland Voices database. This is particularly relevant if TRAP lowering ultimately interacts with START and STRUT, which are traditionally only distinguished by length. It is unclear which vowel lowering would ultimately impact, or whether both would be impacted. Furthermore, Mount Roskill women in conversational speech seem to not produce any onglide for FLEECE. This suggests it is not diphthongized. It would be interesting to be able to compare this result to other conversational speech data to gain further understand of FLEECE diphthongization in Auckland.

Future research might also consider the vowel spaces of other older cohorts of speakers. The Auckland Voices project contains interviews held with older men and women from Mount Roskill, Titirangi, and Papatoetoe. As age effects were not expected to be the most significant factor for this study, it was decided that to limit data preparation time only one group of older speakers would be prepared. The older Titirangi women were selected as sociolinguistic factors suggested they would be the most conservative speakers of New Zealand English in Auckland. Sociolinguistic factors also suggested that the young speakers from Titirangi should sound much like the older group. This however, is not the case, while the older speakers are conservative in

regards to NZE, the young speakers are not. Without data from other older cohorts it is difficult to discuss the sociolinguistic factors contributing to age effects in the data presented here. Analyzing data from all older groups would strengthen the analysis presented here, and it would be interesting to see how older speakers from Papatoetoe and Mount Roskill and responding to language change in Auckland.

Chapter VIII

Conclusion

This thesis set out to answer the question: Is English in Auckland the same or different to regional New Zealand English? To investigate this, an acoustic analysis of the vowel spaces 40 New Zealand English speakers in Auckland, across three suburbs, was undertaken. This analysis was shaped by questions about the impact of gender, age, and diversity in New Zealand. It considered over 12500 vowel tokens.

The acoustic analysis paid special attention to data preparation processes, advocating for careful data preparation. Special consideration was paid to hand checking phoneme and word boundaries, as well as to formant correction. This reduced inaccuracies caused by both automatic software and recording quality. Particular care was also taken in analyzing only stressed vowel tokens at the vowel target(s). This further improved the accuracy of the analysis.

The acoustic analysis provided many important results. The most interesting of which were significant age effects between young and older speakers. For all young speaker groups the New Zealand English TRAP and DRESS vowels are lowering, and the diphthongs FACE, GOAT and PRICE are shifting. There is no evidence, however, for this change amongst the older cohort. A statistical analysis supports these results, finding large differences between old and young speakers, and minimal differences across suburbs and genders. As expected there is evidence of increased variation among Mount Roskill female speakers, however this does not seem to translate to them leading language change in Auckland. In addition, where we expected the young Titirangi speakers to be comparable to the older group, and New Zealand English, they are instead participating in the aforementioned vowel changes as much as speakers from more diverse suburbs. As such there is no evidence of any cohort analyzed leading language change in Auckland.

The changes found amongst young speakers are in themselves interesting. The lowering of TRAP and DRESS for instance marks the reversal of a common sound change process in New Zealand English among Auckland speakers. This is unusual because these vowels are continuing to rise outside of Auckland. It hence seems unlikely that this change in Auckland indicates the beginning of front vowel lowering nation-wide and we might therefore propose it as a feature of an "Auckland" variety of English.

The movement of diphthongs FACE, PRICE, and GOAT is also interesting. The three vowels seem to be involved in a shift, for young speakers, where PRICE is fronting toward FACE, and FACE and GOAT are rising. This change was unexpected and has little previous literature to link to, as such has been challenging to explain. At the very least we can say that young speakers in Auckland are rejecting traditional New Zealand English forms, and pose this change as a possible feature of "Auckland English". Whether this shift is caused by diversity and dialect contact or as a result of increased stigmatization of New Zealand English forms in Auckland is unclear at this stage.

As for why these changes are occurring among young speakers in Auckland, a few possibilities have been discussed. Notably that many of the changes identified seem to be movements away from New Zealand English towards northern hemisphere variations. It is possible that this is the result of the diversity of Auckland and pressure from dialect contact and mixing in a super-diverse city. It is most likely however, that New Zealand English in Auckland is undergoing change as the result of a multitude of different sociolinguistic pressures relating to diversity, identity, and mobility.

There are of course limitations to this study. Most importantly, it largely focuses on read speech, and only considers a small amount of the conversational data available through the Auckland Voices Project. Also because we were expecting suburb and gender to be the most significant factors, we have not considered older speakers from two suburbs. Considering both of these groups would allow deeper examination of some interesting phenomena that could not be commented on in great detail here.

This study has, however, filled a large gap in research on New Zealand English. Until now no studies of the New Zealand English vowel space have considered speakers from New Zealand's largest and most diverse city. This is a major issue, given claims about the uniformity of the New Zealand English vowel space have been made without consideration of one third of its speakers. This study has also suggested that language change processes in diverse cities are not always as clear cut as those found in studies such as Cheshire et al. (2013). While the suggestion is that diversity in a community equates to variation, which results in language change, the actual linguistic scenario can be more complex.

As for the opening question: Is English in Auckland the same or different to regional New Zealand English? The data here seems to suggest that New Zealand English is different in Auckland, particularly for young speakers. If there is an "Auckland English" it would likely be defined by front vowel lowering and diphthong shifts discussed in this paper. It would be irresponsible, however, to claim the existence of "Auckland English" based upon one study. Further studies on Auckland speakers of New Zealand English are necessary make this claim more reliable. However, given previous focus on the lack of regional variation in New Zealand English, even the possibility of "Auckland English" is a major discovery for New Zealand English research.

References

Allan, W. S. & Starks, D. (2000). 'Noone sounds like us? A comparison of New Zealand and other southern hemisphere englishes. In Bell, A. & Kuiper, K. (Eds.) *New Zealand English* (pp. 221-248). Wellington: Victoria University Press.

Al-Tairi, H., Watson, C., & Brown, J. (2016). Secondary Tongue Retraction in Arabic Emphatics: An Acoustic Study. Proceedings from SST2016: *16th Australasian International Conference on Speech Science and Technology.* Parramatta, Australia.

Auckland Council. (2014a). *Waitākere Ranges local board profile - Initial results from the 2013 Census.* Retrieved May 10, 2018, from http://temp.aucklandcouncil.govt.nz /SiteCollectionDocuments/aboutcouncil/localboards/Waitakererangeslocalboard/wait% C4%81kererangeslocalboardcensusprofile.pdf

Auckland Council. (2014b). *Ōtara-Papatoetoe local board profile - Initial results from the 2013 Census.* Retrieved May 10, 2018, from http://temp.aucklandcouncil.govt. nz/SiteCollectionDocuments/aboutcouncil/localboards/otarapapatoetoelocalboard/otar apapatoetoelocalboardcensusprofile.pdf

Auckland Council. (2014c). *Puketāpapa local board profile - Initial results from the 2013 Census.* Retrieved May 10, 2018, from http://temp.aucklandcouncil.govt.nz/SiteColl ectionDocuments/aboutcouncil/localboards/Puketapapalocalboard/Puketapapalocalbo ardcensusprofile.pdf

Auckland Council. (2017). *Puketāpapa Local Board Plan 2017.* Retrieved June 28, 2018, from https://www.aucklandcouncil.govt.nz/about-auckland-council/how-auckland-council- works/local-boards/all-local-boards/Documents/Puketapapa-local-board-plan-2017.pdf

Audacity Team (2017). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.1 Retrieved December 20th 2017 from https://audacityteam.org/

Australian Bureau of Statistics. (2014). *Australian social trends, 2014: where do migrants live?* Retrieved June 25, 2018, from http://www.abs.gov.au/ausstats/abs%40.nsf/Lookup/41 02.0main%2Bfeatures102014.

Bartlett, C. (1992). Regional variation in New Zealand English: The case of Southland. *New Zealand English Newsletter, 6*, 5-15. University of Canterbury.

Bauer, L. (1979). The second Great Vowel Shift. *Journal of the International Phonetic Association, 9*(2), 57-66.

Bauer, L. (1992). The second Great Vowel Shift revisited. *English World-Wide, 13*(2), 253-268.

Bauer, L. (1994). English in New Zealand. In Burchfield, R (Ed.), *The Cambridge History of the English Language: Volume 5.* Cambridge: Cambridge University Press.

Bell, A. (1997). The phonetics of fish and chips in New Zealand: Marking national and ethnic identities. *English World-Wide, 18*(2), 243-270.

Bell, A. (2000). Māori and Pākehā English: a case study. In Bell, A. & Kuiper, K. (Eds.), *New Zealand English* (pp. 221-248). Wellington: Victoria University Press.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Britain, D. (2008). When is a change not a change? A case study on the dialect origins of New Zealand English. *Language Variation and Change*, *20*(2), 187-223.

Bybee, Joan. (2003). *Phonology and language use*. Cambridge: Cambridge University Press.

Cheshire, J., Fox, S., Kerswill, P. & Torgerson, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics, 15*(2), 151–196.

Cheshire, J., Fox, S., Kerswill, P. & Torgerson, E. (2013). Language contact and language change in the multicultural metropolis. *Revue Française de Linguistique Appliquée, 17*(2), 63-76.

Cox, F., & Palethorpe, S. (2001). The changing face of Australian English vowels. In Blair, D. & Collins, P. (Eds.), *Varieties of English around the world: English in Australia*, (pp. 17-44). Amsterdam: John Benjamins Publishing Company.

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341-350.

Cox, F., & Palethorpe, S. (2008). Reversal of short front vowel raising in Australian English. In *Ninth Annual Conference of the International Speech Communication Association*, 342-355.

Disner, S. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America*, *67*(1), 253-261.

Easton, A., & Bauer, L. (2000). An acoustic study of the vowels of New Zealand English. *Australian Journal of Linguistics*, *20*(2), 93-117.

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America*, *140*(1), 576-581.

Gargesh, R. (2008). Indian English: Phonology. *Varieties of English*, *4*, 231-243.

Gibson, A. & Bell, A. (2010). Performing Pasifika English in New Zealand. *English World-Wide, 31*(3), 231-251.

Google Maps. (2018). *Auckland, New Zealand*. Retrieved from https://www.google.co.nz /maps/@-36.9399533,174.7396728,12z

Gordon, E. (1998). The origins of New Zealand speech: The limits of recovering historical information from written records. *English World-Wide, 19(1),* 61-85.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: its origins and evolution*. New York: Cambridge University Press.

Gordon, E. & Deverson, T. (1998). *New Zealand English and English in New Zealand.* Auckland: New House Publishers.

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.

Hardcastle, W. & Barry, W. (1989). Articulatory and perceptual factors in/l/vocalisations in English. *Journal of the International Phonetic Association*, *15*(2), 3-17.

Harrington, J. (2010). *Phonetic Analysis of Speech Corpora.* United Kingdom: Wiley-Blackwell.

Harrington, J. & Cassidy, S. (1999). *Techniques in Speech Acoustics.* Dordrecht: Kluwer Academic Publishers.

Hay, J., Maclagan, M. & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press.

Holmes, J. & Bell, A. (1992). On shear markets and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English. *Language Variation and Change, 4,* 251-273.

Holmes, J. & Boyce, M. (1991). *Variation and change in New Zealand English: a social dialect study.* Wellington: Victoria University of Wellingotn.

Horvath, B.M. & Horvath, R.J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalisation in New Zealand and Australian English. *Language Variation and Change, 13*, 37-57.

James, J., Tian, L. & Watson, C.I. (2018). An Open Source Emotional Speech Corpus for Human Robot Interaction Applications. In *The Proceedings of Interspeech 2018*, Hyderabad India, Sept 2-6 2018.

Johnson, W. & Britain, D. (2007). L-vocalisation as a natural phenomenon: explorations in sociophonology. *Language Sciences*, *29*(2-3), 294-315.

Kisler, T., Reichel U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45*, 326–347.

Labov, W. (1972). Some principles of linguistic methodology. *Language in society*, *1*(1), 97-120.

Labov, W. (1994). Principles of language change: Internal factors. Oxford: Blackwell.

Lass, R. (1990). Where do extraterritorial Englishes come from? In Adamson, S., Law, V., Vincent, N. & Wright, S. (Eds.), *Papers from the 5th international conference of English historical linguistics* (pp. 245-280)*.* Amsterdam: John Benjamins Publishing Company.

Maclagan, M. (1982). An acoustic study of New Zealand vowels. *New Zealand Speech Therapists' Journal*, *37*(1), 20-26.

Maclagan, M. A., Gordon, E., & Lewis, G. (1999). Women and sound change: Conservative and innovative behavior by the same speakers. *Language variation and change*, *11*(1), 19-41.

Maclagan, M., King, J. & Gillon, G. (2008). Māori English. *Clinical Linguistics & Phonetics, 22*(8), 658-670.

Maclagan, M. & Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change, 8,* 125-147.

Maclagan, M. & Gordon, E. (2004) The story of New Zealand English: What the ONZE Project tells us. *Australian Journal of Linguistics, 24*(1), 41-56.

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.

Maclagan, M., Watson, C., Harlow, R., King, J. & Keegan, P. (2017) Investigating the sound change in the New Zealand English nurse vowel /ɜ:/. *Australian Journal of Linguistics, 37*(4), 465-485.

Meyerhoff, M. (2017). *Community dependencies Connections and discontinuities in Auckland City*. Key note lecture at the New Zealand Linguistic Society Conference, Auckland, 23-24 November 2017.

Meyerhoff, M., Schleef, E., & MacKenzie, L. (2015). *Doing sociolinguistics: A practical guide to data collection and analysis*. Oxford: Routledge.

New Zealand Parliament. (2008). Immigration chronology: selected events 1840-2008. Retrieved 20 June, 2018, from https://www.parliament.nz/en/pb/research-papers/document/00PLSocRP08011/immigration-chronology-selected-events-1840-2008#_Toc196101768

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. (2018). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-137, https://CRAN.R-project.org/package=nlme.

R Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/

Rosen, S. & Howell, P. (1991). *Signals and systems for speech and hearing.* London: Academic Press.

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Sneddon, R. (2015). Multicultural London English/Multicultural Paris French.. Implications for the classroom. *Éducation et Sociétés Plurilingues*, *39*, 79-88.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291-311.

Starks, D., & Reffell, H. (2006). Reading 'TH': Vernacular variants in Pasifika Englishes in South Auckland. *Journal of Sociolinguistics*, 10(3), 382-392.

Starks, D., Christie, J., & Thompson, L. (2007). Niuean English: Initial insights into an emerging variety. *English world-wide*, 28(2), 133-146.

Starks, D., Gibson, A. & Bell, A. (2015). Pasifika Englishes in New Zealand. In Williams, J. P., Schneider, E.W., Trudgill, P. & Schreier, D. (Eds.), *Further studies in the lesser known varieties of English* (pp. 288-304).

Statistics NZ. (2013a). 2013 QuickStats about a place: Auckland Region - Population and dwellings. [Data file]. Retrieved 1 May, 2018, from http://archive.stats.govt.nz/ Census/2013-census/profile-and-summary-reports/quickstats-about-a-place.aspx?request_v alue=13170&tabname=Populationanddwellings

Statistics NZ. (2013b). 2013 QuickStats about a place: Auckland Region - Cultural Diversity. [Data file]. Retrieved 1 May, 2018, from http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-about-a-place.aspx?request_value=131 70&reportid=10&tabname=Culturaldiversity

Statistics NZ. (2013c). 2013 Census QuickStats about a place: Otara-Papatoetoe Local Board Area. [Data file]. Retrieved May 10, 2018, from http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quick stats-about-a-place.aspx?url=%2FCensus%2F2013-census%2Fprofile-and-summary-rep

orts%2Fquickstats-about-a-place.aspx&request_value=13626&tabname=&p=y&
printall=true

Statistics NZ. (n.d.*a*). Statsmaps: 2013 Census map - ethnicity. Retrieved May 12, 2018, from
http://archive.stats.govt.nz/StatsMaps/Home/People%20and%20households/2013-
census-map-ethnicity-as-a-percentage-of-total-population.aspx

Statistics NZ. (n.d.*b*). Mapping trends in the Auckland Region. Retrieved June 30, 2018 from
http://archive.stats.govt.nz/browse_for_stats/population/census_counts/census-
snapshot-cultural-diversity.aspx

Sydney Speaks Project. (n.d.). Sydney Speaks: Language variation and change in a diverse
society. Retrieved on 15 June, 2018 from http://www.dynamicsoflanguage.edu.au/
sydney-speaks/

Szakay, A. (2007). *Identifying Māori English and Pākehā English from suprasegmental cues: A
study based on speech resynthesis.* (Master's Thesis, University of Canterbury).

Thompson, L., Watson, C. I., & Starks, D. (2009). Finding one's own vowel space. In *World
Englishes - Problems, Properties and Prospects: Selected Papers from the 13th IAWE
Conference* (Vol. 40). John Benjamins Publishing.

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.

Trudgill, P., Gordon, E., Lewis, G. & Maclagan, M. (2000) Determinism in new-dialect formation
and the genesis of New Zealand English. *Journal of Linguistics, 36*(2), 299-318.

Turner, G. W. (1967). Samuel McBurney's newspaper article on colonial pronunciation. *Journal of the Australasian Universities Language and Literature Association*, 27(1), 81-85.

Wall, A. (1938). *New Zealand English: How it should be spoken.* New Zealand: Whitcombe and Tombs limited.

Warren, P. (2017). Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association,* 1-26.

Warren, P. & Bauer, P. (2008). Māori English: Phonology. In Burridge, K., & Kortmann, B. (Eds.). *The Pacific and Australasia,* (pp. 77-88)*.* Berlin: Mouton de Gruyter.

Watson, C., Harrington, J. & Evans, Z. (1998). An acoustic comparison between New Zealand and Australian English vowels. *Journal of Linguistics, 18*(2), 187-287.

Watson, C., Maclagan, M. & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change, 12,* 51-68.

Watson, C., Maclagan, M., King, J., Harlow, R. & Keegan, P. (2016). Sound change in Māori and the influence of New Zealand English. *Journal of the International Phonetic Association, 46(2)*, 185-218.

Watson, C., Palethorpe, S. & Harrington, J. (2004) Capturing the vowel change in New Zealand English over a thirty year period via a diachronic study. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology* (pp. 201-206).

Wells, J. C. (1982a). Accents of English 1: An Introduction. Cambridge: Cambridge University Press.

Wells, J. C. (1982b). Accents of English 3: Beyond the British Isles. Cambridge: Cambridge University Press.

Winkelmann, R. & Raess, G. (2014). Introducing a web application for labeling, visualizing speech and correcting derived speech signals. In *LREC* (pp. 4129-4133).

Winkelmann, R., Jaensch, K., Cassidy, S. & Harrington, J. (2018). emuR: Main Package of the EMU Speech Database Management SystemR package version 1.0.0.

Woods, N. J. (1997). The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change. *Journal of Sociolinguistics*, *1*(1), 95-125.

Appendix A

Read Speech Passage

Read Speech

1. Last summer we went touring around the East Cape Area in our old van.
2. We towed our boat behind us.
3. One day we stopped at a small bay for a bite to eat.
4. I checked the water, topped up the oil, and followed the family down to the beach.
5. The sea is just great for swimming said Carl.
6. During lunch I decided to sail around a little island about half a mile offshore.
7. Carl was sleepy and the girls were playing in a pool with their dolls so I went alone.
8. It took about an hour with a fairly strong wind to get all the way around the island.
9. By then, I could see the shore again and I started to feel rather pleased with myself.
10. That was when things began to go wrong.
11. The breeze which had blown me out so fast was getting stronger.
12. I pulled in the sail tight, but the boat was so light that I was being blown out to sea.
13. At first there had been plenty of people around.
14. But now the sea was almost empty and there was noone within hearing distance.
15. I howled till my voice gave up, but it was a waste of time.
16. I was wearing only shorts and a t-shirt and the wind was cruel.
17. The swell grew bigger but I tried not to panic.
18. I said to myself "Carl knows I'm out here, he'll get help".
19. I decided to take down the sail and row, but still nothing happened
20. I made scarcely any progress.
21. The wind got even fiercer and the sky grew dark.
22. I was in real trouble and I felt awful, I began to pray.
23. I don't usually believe in miracles but that's how I felt about what happened next.
24. A motorboat appeared around the island heading straight towards me.
25. One of the men in it threw me a rope.
26. "Catch Pal!" he called.
27. I tied it to a wire hoop on the hull, and they towed me back to shore.
28. A woman had spotted me through her binoculars.
29. She saw I was in trouble and sent the boat out to get me.
30. And where was Carl?
31. Well he'd driven around to the next bay.
32. He thought I must've sailed in there by mistake.
33. Lucky I hadn't waited for him to rescue me.

Appendix B
Formant Track Errors

Section 4.1.3 discussed automatic formant tracking and the necessity of hand checking formants. This Appendix provides some extra examples of formant track errors found in our database. This is so the reader can better understand how noisy recordings impact formant calculation, and the necessity of careful data preparation.
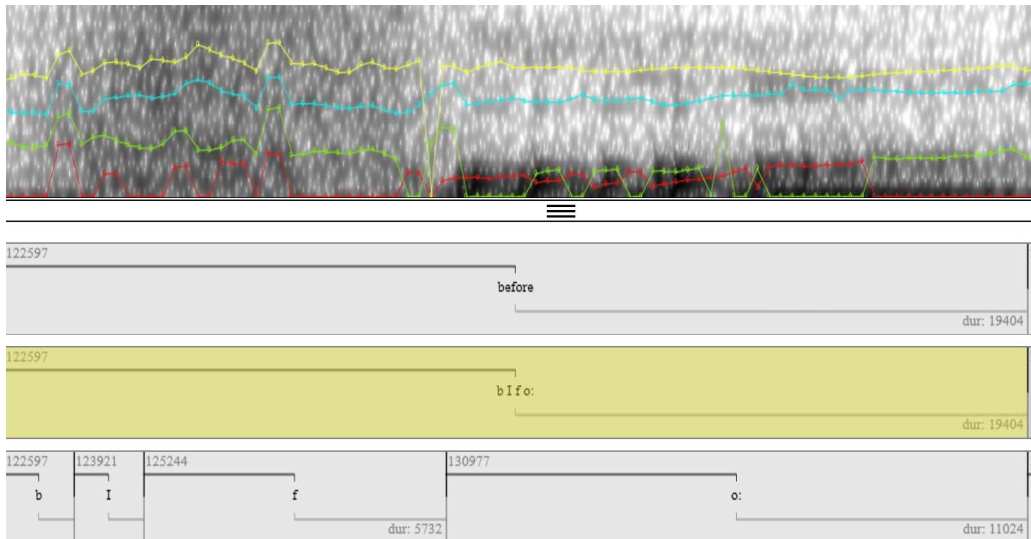

Figure B1. Example of formant track error in the word 'before': F1 (*red*) and F2 (*green*) switch during THOUGHT vowel production.
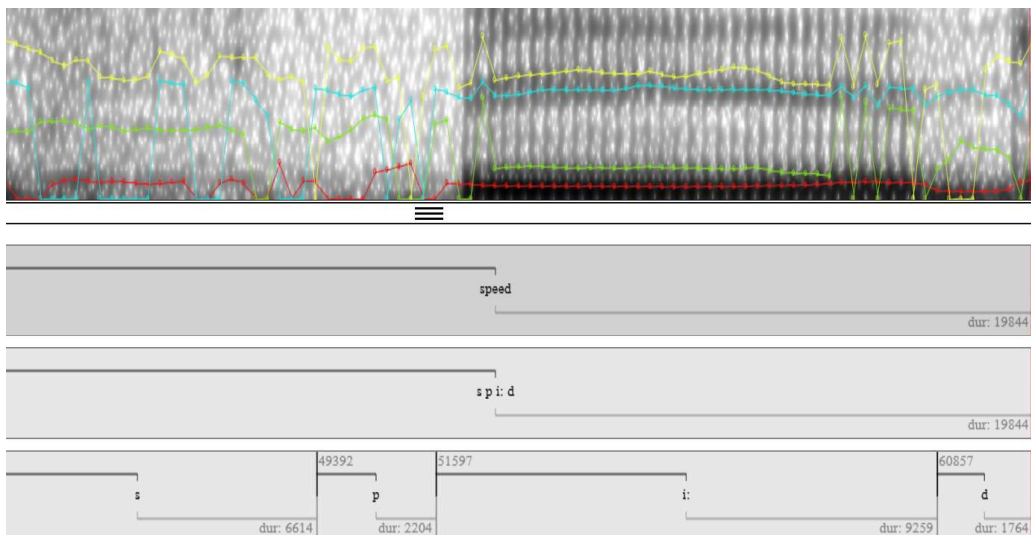

Figure B2. Example of formant track error in the word 'speed': During the production of FLEECE, F3 (*blue*) is automatically tracked along actual second formant; F2 (*green*) is tracked along vowel harmonic rather than formant.

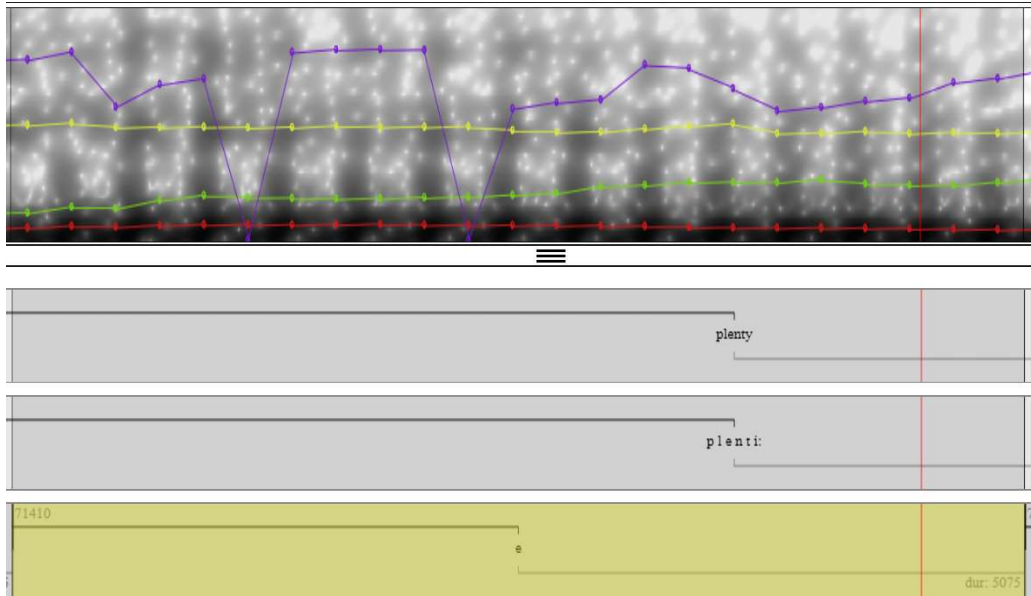*Figure B3.* Example of formant track error in the word 'plenty': During the production of a DRESS vowel, F3 (*yellow*) automatically tracked along actual second formant; F2 (*green*) tracked along vowel harmonic rather than formant.
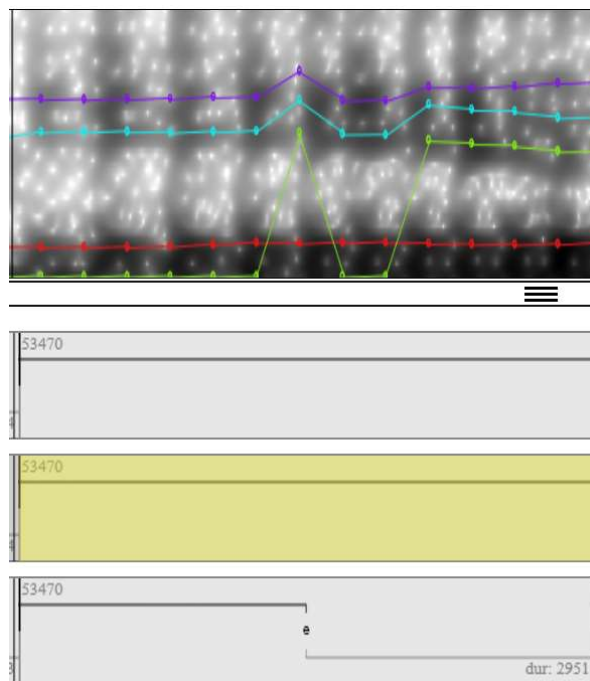

Figure B4. Example of formant track error in the word 'empty': Formant tracker cannot locate F2 (green) and hence is tracking along zero Hz during DRESS vowel production.
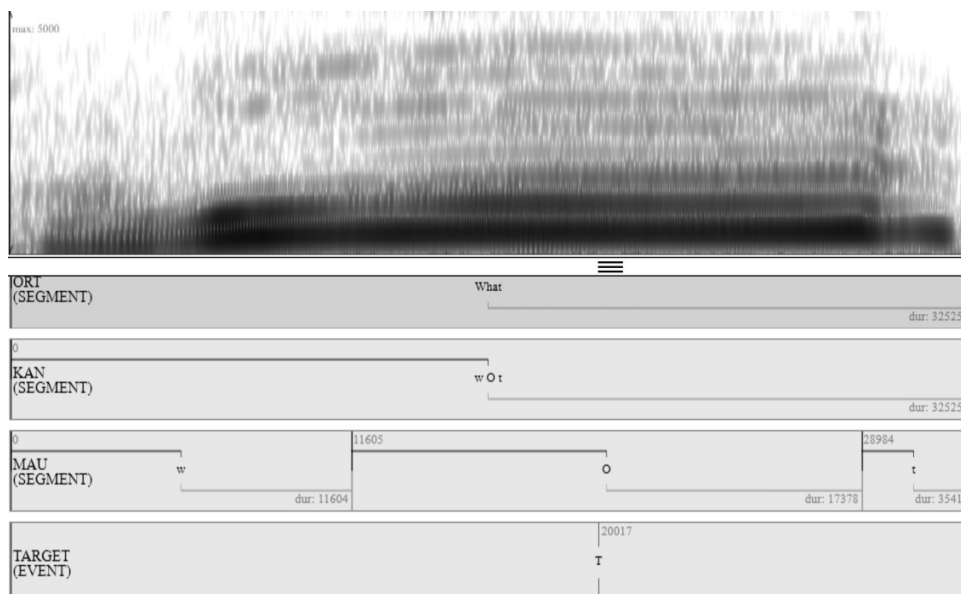
Figure B5. Example showing the Impact of 'clipping' on a spectrogram during the word 'what'. Speaker overclocked the microphone while producing the LOT vowel. The Token had to be removed as formants were unidentifiable under excess spectral noise.
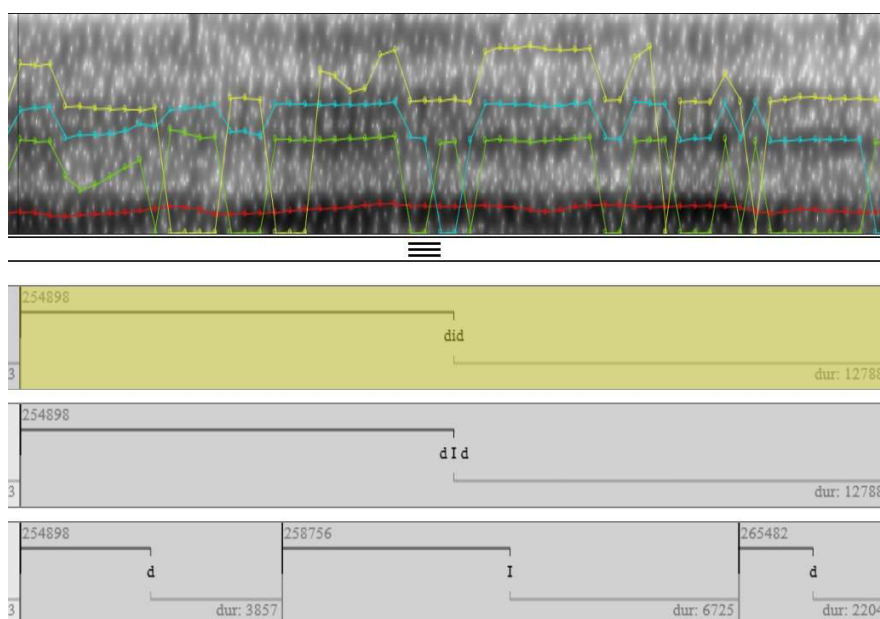
Figure B6. Example of formant track error in the word 'empty': This shows the impact of background noise on the spectrogram and formant calculation in the word 'did'. Recording contained background noise from cicadas in nearby tree: Multiple tracking errors and barely visible formants.
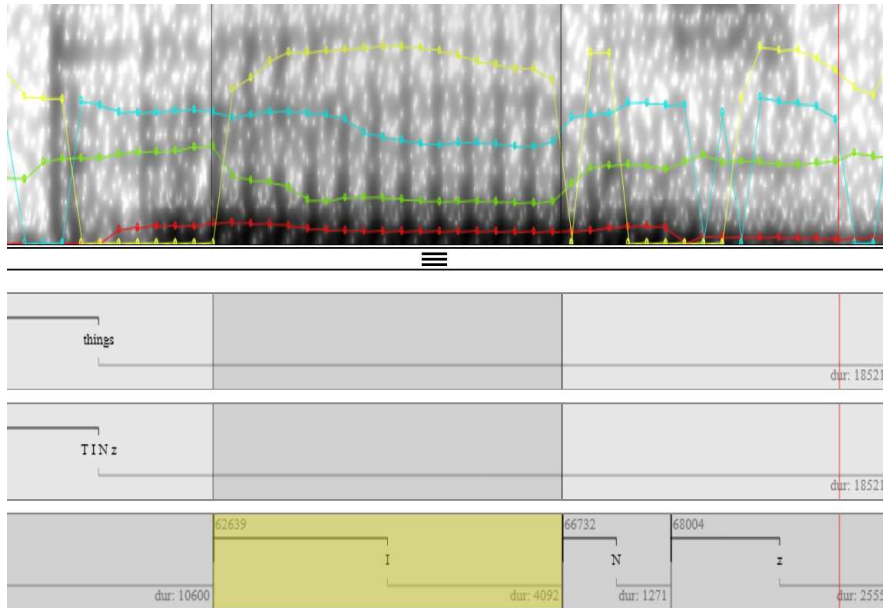
Figure B7. Example of formant track errors cause by recording noise during the word 'things': F2 and F3 are mistracked and formants are not visible on spectrogram. F1 is mistracked along F0. Formants were located and corrected using spectral slices from corresponding *PRAAT* TextGrid.

Appendix C

Post vocalic L impact: Read speech

Figure C1. Titirangi Male Speakers – Effects of post vocalic /l/ on individual vowel tokens (*left*), and centroids (*right*).

Figure C2. Titirangi Female Speakers – Effects of post vocalic /l/ on individual vowel tokens (*left*), and centroids (*right*).
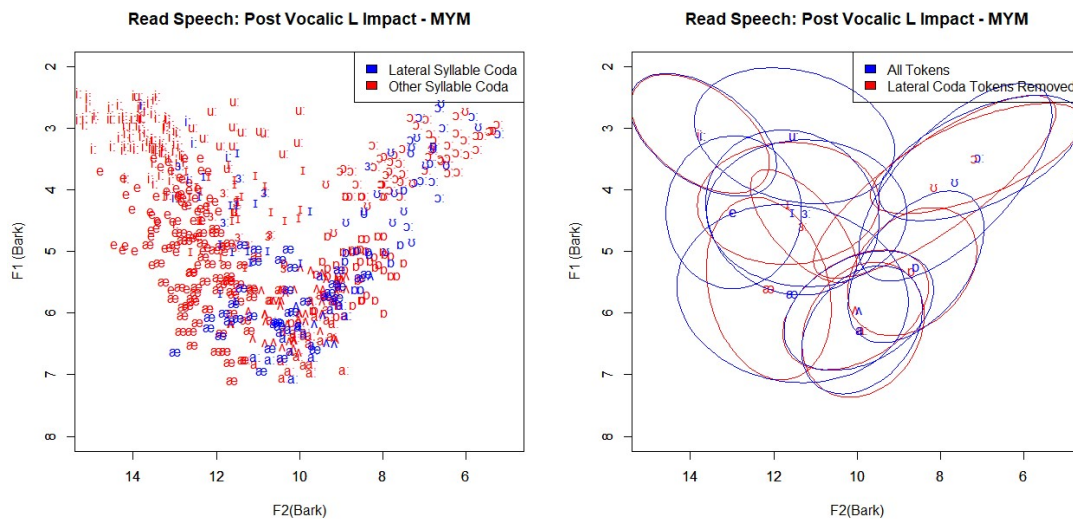
Figure C3. Mount Roskill Male Speakers – Effects of post vocalic /l/ on individual vowel tokens (left), and centroids (right).
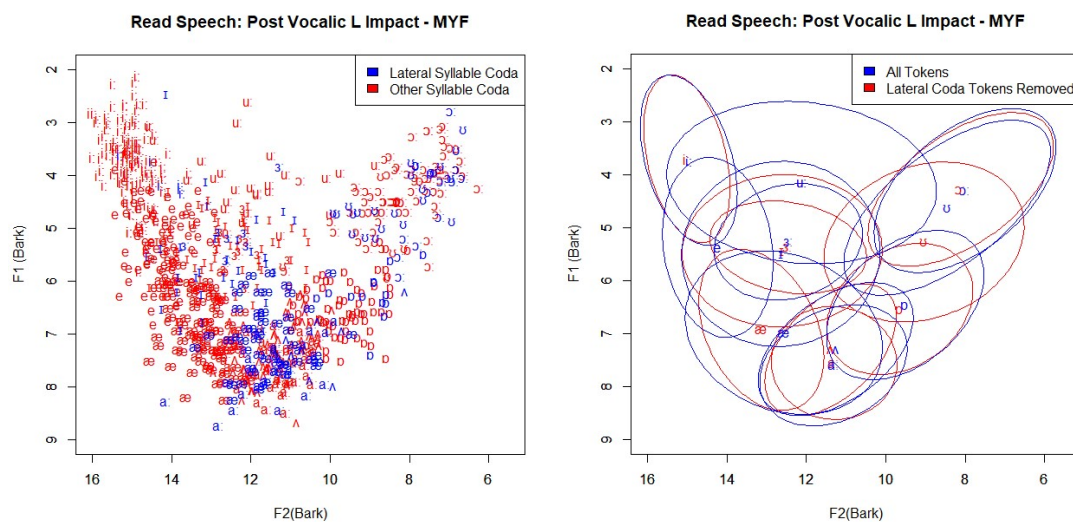
Figure C4. Mount Roskill Female Speakers – Effects of post vocalic /l/ on individual vowel tokens (left), and centroids (right).
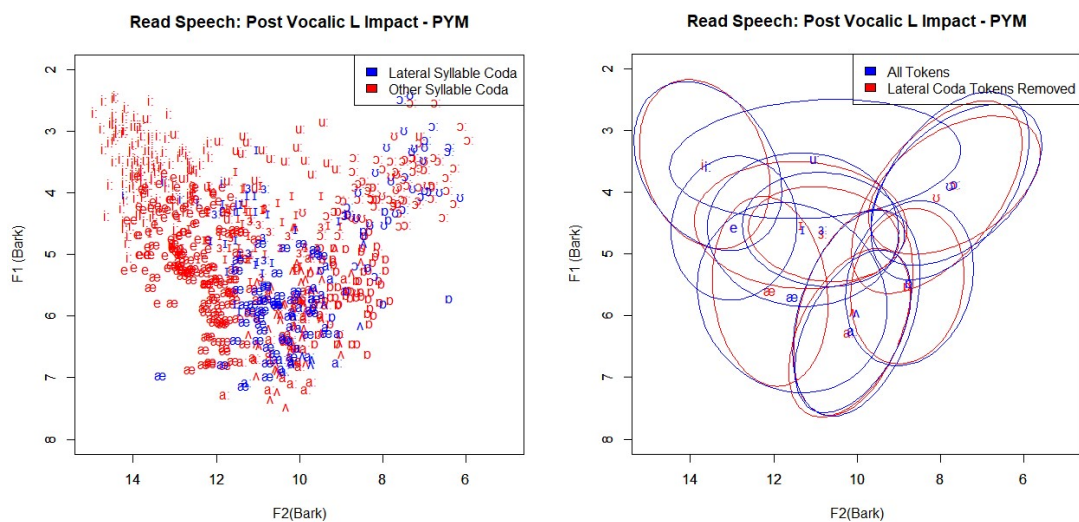
Figure C5. Papatoetoe Male Speakers – Effects of post vocalic /l/ on individual vowel tokens (left), and centroids (right).
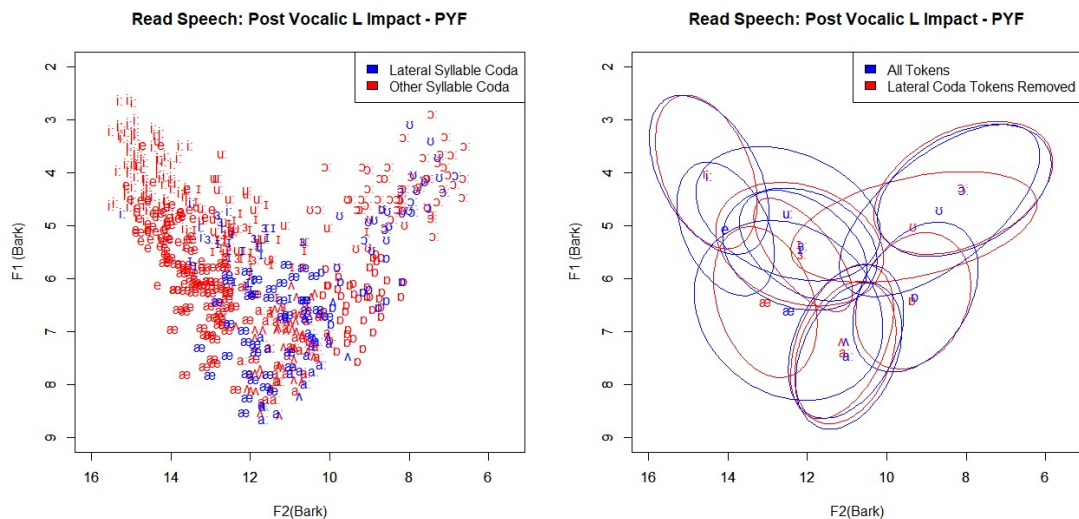
Figure C6. Papatoetoe Female Speakers – Effects of post vocalic /l/ on individual vowel tokens (left), and centroids (right).
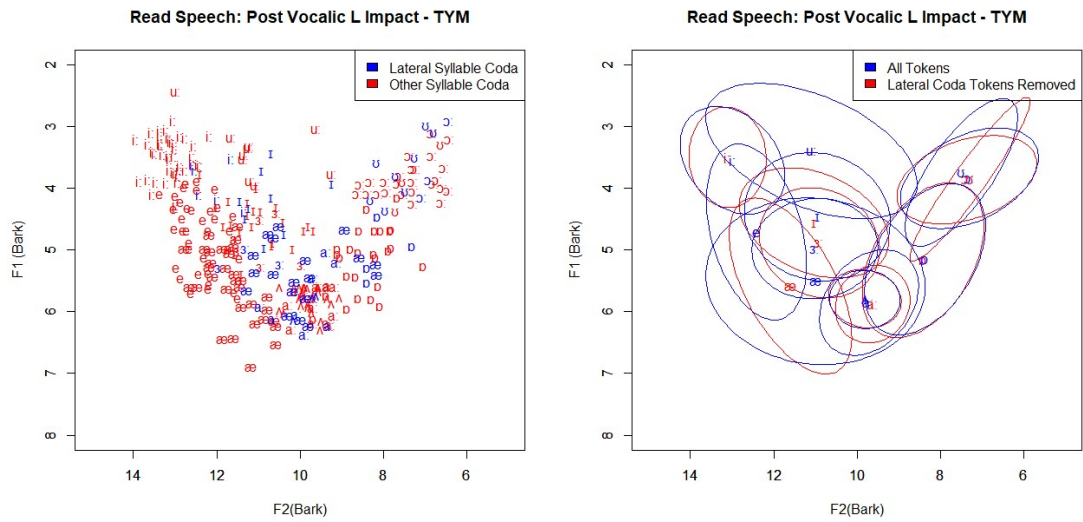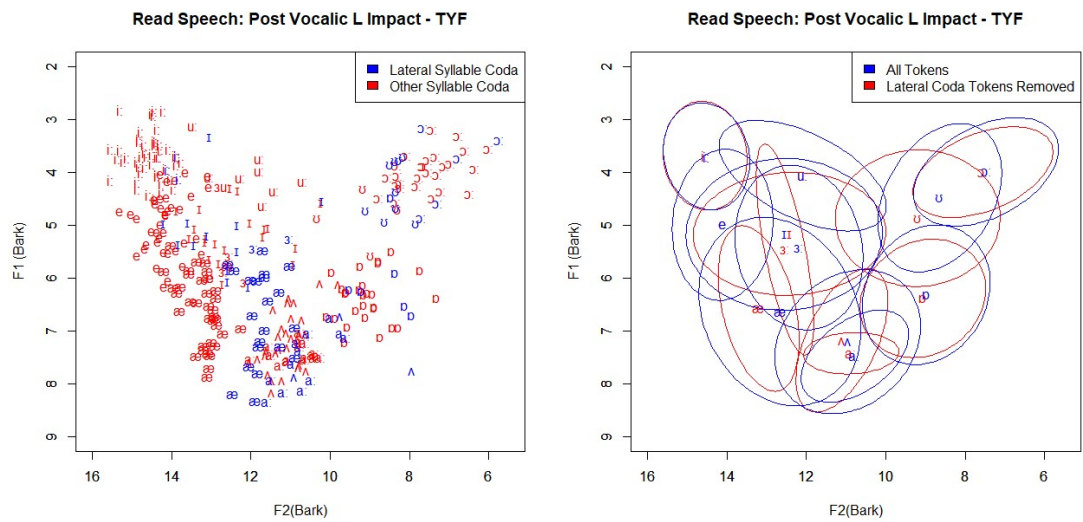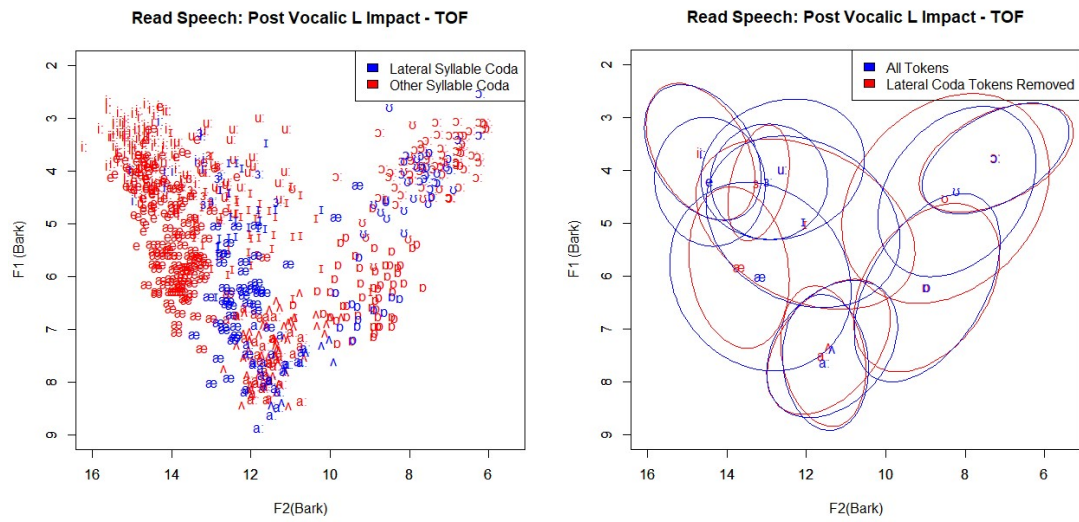
Figure C7. Older Titirangi Speakers – Effects of post vocalic /l/ on individual vowel tokens (left), and centroids (right).

Appendix D

Token numbers: All Vowels

(D1)

Read Speech: Titirangi Male Speakers (n=3)
Total Stressed Monophthongs (All Tokens) & Total Stressed Monophthongs: Tokens With /l/ Coda Removed (No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	51	45
GOOSE	10	10
DRESS	37	37
KIT	31	22
NURSE	9	6
TRAP	51	32
STRUT	28	25
START	23	15
THOUGHT	28	24
LOT	30	26
FOOT	11	3

(D2)

Read Speech: Mount Roskill Male Speakers (n=6)
Total Stressed Monophthongs (All Tokens) & Total Stressed Monophthongs: Tokens With /l/ Coda Removed(No /l/ Coda).

Monophthong	All Tokens	No /l/ Coda
FLEECE	95	89
GOOSE	16	16
DRESS	68	68
KIT	46	31
NURSE	16	11
TRAP	95	61
STRUT	55	49
START	56	36
THOUGHT	66	50
LOT	62	51
FOOT	25	7

(D3)

Read Speech: Papatoetoe Male Speakers (n=7)
Total Stressed Monophthongs (All Tokens) & Total Stressed Monophthongs: Tokens With /l/ Coda Removed (No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	130	122
GOOSE	27	27
DRESS	93	93
KIT	74	51
NURSE	19	12
TRAP	126	79
STRUT	70	63
START	69	42
THOUGHT	79	62
LOT	74	63
FOOT	31	12

(D4)

Read Speech: Titirangi Female Speakers (n=3)
Total Stressed Monophthongs (All Tokens) & Total Stressed Monophthongs: Tokens With /l/ Coda Removed(No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	52	49
GOOSE	9	9
DRESS	40	40
KIT	32	20
NURSE	6	4
TRAP	58	37
STRUT	32	29
START	30	18
THOUGHT	32	24
LOT	33	27
FOOT	13	4

Read Speech: Mount Roskill Female Speakers n=(8)
Total Stressed Monophthongs (All Tokens) & Total Stressed
Monophthongs: Tokens With /l/ Coda Removed (No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	146	138
GOOSE	28	28
DRESS	99	99
KIT	83	56
NURSE	21	14
TRAP	142	89
STRUT	75	67
START	71	48
THOUGHT	90	68
LOT	82	68
FOOT	30	7

Read Speech: Papatoetoe Female Speakers (n=6)
Total Stressed Monophthongs (All Tokens) & Total Stressed
Monophthongs: Tokens With /l/ Coda Removed (No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	102	96
GOOSE	29	29
DRESS	73	73
KIT	57	39
NURSE	16	10
TRAP	101	61
STRUT	55	49
START	53	33
THOUGHT	59	73
LOT	56	45
FOOT	25	8

(D7)

Read Speech: Older Titirangi Female Speakers (n=7)
Total Stressed Monophthongs (All Tokens) & Total Stressed Monophthongs: Tokens With /l/ Coda Removed (No /l/ Coda).

Vowel	All Tokens	No /l/ Coda
FLEECE	137	127
GOOSE	31	31
DRESS	91	91
KIT	73	50
NURSE	17	10
TRAP	127	79
STRUT	65	58
START	65	39
THOUGHT	80	59
LOT	68	54
FOOT	29	10

(D8)

Conversational Speech Monophthong Totals
No. Vowel Tokens Papatoetoe Male Speakers (n=7) (Papatoetoe)
No. Vowel Tokens Mount Roskill Female Speakers (n=6) (Mount Roskill)

Vowel	Papatoetoe Male	Roskill Female
FLEECE	269	229
GOOSE	157	114
DRESS	335	220
KIT	273	311
NURSE	113	229
TRAP	304	242
STRUT	233	222
START	119	98
THOUGHT	152	110
LOT	210	177
FOOT	87	63

(D9)

Read Speech: *Diphthong Token No. by Vowel & Suburb/Gender*

Suburb/Gender	FACE	GOAT	MOUTH	NEAR	SQUARE	PRICE	CHOICE
Titirangi Male	46	35	30	10	13	42	7
Titirangi Female	49	37	27	9	17	43	6
Mount Roskill Male	87	72	35	17	23	79	9
Mount Roskill Female	127	103	82	29	37	109	16
Papatoetoe Male	111	92	83	18	35	82	15
Papatoetoe Female	97	80	64	19	32	74	12
Titirangi Female - Older	119	106	82	23	30	102	12

(D10)

Conversational Speech: *Diphthong Token No. by Vowel and Suburb.*

Suburb	FACE	GOAT	MOUTH	NEAR	SQUARE	PRICE	CHOICE
Mount Roskill Female	186	139	63	30	54	219	13
Papatoetoe Male	248	135	88	10	81	229	7

Appendix E
Monophthong t-test results

Table E1. *Null model*Suburb: t-test results for significant vowels.*

Suburb differences: NURSE

	Value	Std.Error	DF	t-value	p-value
typeF1	4.94378	0.08407	85	58.80371	1.2410e-70
typeF2	11.04592	0.14619	85	75.55829	1.0295e-79
subP	0.13381	0.12202	30	1.09664	0.28152
subT	0.34101	0.14489	30	2.35353	0.02534
typeF2:subP	-0.58158	0.26317	85	-2.20985	0.02980
typeF2:subT	-0.45962	0.32101	85	-1.43182	0.15586

Suburb differences: FLEECE

	Value	Std.Error	DF	t-value	p-value
typeF1	3.38172	0.08744	1054	38.67543	2.07e-204
typeF2	13.23940	0.10787	1054	122.73907	0.00e+00
subP	0.35459	0.12583	31	2.81808	0.00834
subT	0.21689	0.15929	31	1.36160	0.18314
typeF2:subP	-0.60613	0.23805	1054	-2.54619	0.01103
typeF2:subT	-0.76763	0.30146	1054	-2.54635	0.01103

Table E2. *Null model*Gender: t-test results for significant vowels.*

Gender differences: GOOSE

	Value	Std.Error	DF	t-value	p-value
typeF1	4.08409141	0.0996578	203	40.9811409	3.97e-100
typeF2	10.8763078	0.1723387	203	63.1100646	2.28e-135
sexM	-0.5656442	0.1459013	32	-3.8768971	0.0004943
typeF2:sexM	0.46658623	0.2752183	203	1.69533162	0.0915457

Table E3. *Null model*Gender&Suburb: t-test results for significant vowels.*

Gender*Suburb Differences: TRAP

	Value	Std.Error	DF	t-value	p-value
typeF1	3.8977	0.11792	199	33.05326	9.32e-83
typeF2	10.7879	0.25816	199	41.78773	1.81e-100
subP	0.5276	0.17201	28	3.06788	0.00475
subT	-0.11922	0.23234	28	-0.51315	0.61187
sexM	-0.58652	0.19015	28	-3.0846	0.00455
typeF2:subP	-0.24238	0.38591	199	-0.62809	0.53067
typeF2:subT	0.076929	0.53303	199	0.14432	0.88539
typeF2:sexM	1.00578	0.43678	199	2.30271	0.02233
subP:sexM	-0.23229	0.25921	28	-0.89612	0.37783
subT:sexM	0.37291	0.33802	28	1.10325	0.27931
typeF2:subP:sexM	-0.72428	0.58818	199	-1.23138	0.21963
typeF2:subT:sexM	-0.83994	0.77526	199	-1.08344	0.27992

Gender*Suburb Differences: FLEECE

	Value	Std.Error	DF	t-value	p-value
typeF1	3.48703	0.11017	1051	31.65249	5.64e-155
typeF2	13.27730	0.13588	1051	97.71368	0.00e+00
subP	0.26499	0.16859	28	1.57179	0.12723
subT	-0.02311	0.21112	28	-0.10946	0.91362
sexM	-0.24918	0.16911	28	-1.47347	0.15178
typeF2:subP	-0.71690	0.32050	1051	-2.23681	0.02551
typeF2:subT	-0.38606	0.40154	1051	-0.96144	0.33655
typeF2:sexM	0.16023	0.32102	1051	0.49912	0.61780
subP:sexM	0.21992	0.24230	28	0.90765	0.37181
subT:sexM	0.51821	0.30555	28	1.69601	0.10098
typeF2:subP:sexM	0.17052	0.46039	1051	0.37038	0.71118
typeF2:subT:sexM	-0.78802	0.58090	1051	-1.35656	0.17521

Gender*Suburb Differences: GOOSE

	Value	Std.Error	DF	t-value	p-value
typeF1	3.89765	0.11792	199	33.05326	9.32e-83
typeF2	10.78790	0.25816	199	41.78773	1.81e-100
subP	0.52770	0.17201	28	3.06788	0.00475
subT	-0.11922	0.23234	28	-0.51315	0.61187
sexM	-0.58653	0.19015	28	-3.08460	0.00455
typeF2:subP	-0.24238	0.38591	199	-0.62809	0.53067
typeF2:subT	0.07693	0.53303	199	0.14432	0.88539
typeF2:sexM	1.00578	0.43678	199	2.30271	0.02233
subP:sexM	-0.23229	0.25921	28	-0.89612	0.37783
subT:sexM	0.37292	0.33802	28	1.10325	0.27931
typeF2:subP:sexM	-0.72428	0.58818	199	-1.23138	0.21963
typeF2:subT:sexM	-0.83995	0.77526	199	-1.08344	0.27992

Gender*Suburb Differences: THOUGHT

	Value	Std.Error	DF	t-value	p-value
typeF1	4.00849	0.06951	519	57.66561	7.96e-228
typeF2	7.41268	0.14802	519	50.07765	7.19e-201
subP	0.00503	0.10743	28	0.04684	0.96297
subT	-0.23005	0.13460	28	-1.70908	0.09850
sexM	-0.32842	0.10662	28	-3.08037	0.00460
typeF2:subP	-0.08794	0.19033	519	-0.46201	0.64427
typeF2:subT	-0.24221	0.23851	519	-1.01553	0.31033
typeF2:sexM	-0.03401	0.18915	519	-0.17979	0.85739
subP:sexM	0.48001	0.15330	28	3.13129	0.00405
subT:sexM	0.61058	0.19257	28	3.17076	0.00367
typeF2:subP:sexM	0.13214	0.27188	519	0.48601	0.62717
typeF2:subT:sexM	0.01609	0.34184	519	0.04706	0.96248

Gender*Suburb Differences: KIT

	Value	Std.Error	DF	t-value	p-value
typeF1	5.01029	0.07468	417	67.08904	1.49e-225
typeF2	11.17793	0.11315	417	98.78641	2.10e-291
subP	-0.07364	0.11583	28	-0.63575	0.53010
subT	-0.24060	0.14498	28	-1.65953	0.10817
sexM	-0.50889	0.12250	28	-4.15421	0.00028
typeF2:subP	-0.28285	0.21279	417	-1.32926	0.18449
typeF2:subT	0.08305	0.26594	417	0.31228	0.75498
typeF2:sexM	0.65350	0.22741	417	2.87362	0.00427
subP:sexM	0.44989	0.17038	28	2.64047	0.01338
subT:sexM	0.53935	0.20480	28	2.63356	0.01360
typeF2:subP:sexM	-0.33881	0.31446	417	-1.07743	0.28191
typeF2:subT:sexM	-1.00963	0.37270	417	-2.70892	0.00703

Appendix F

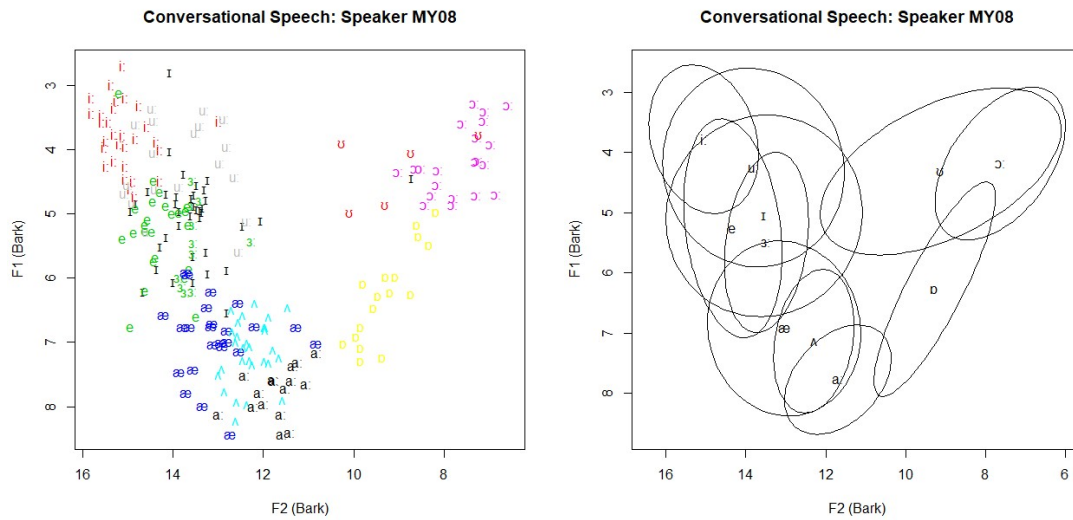Conversational speech: Individual speaker plots

Figure F1. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY02 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.

Figure F2. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY03 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.

Figure F3. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY04 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.
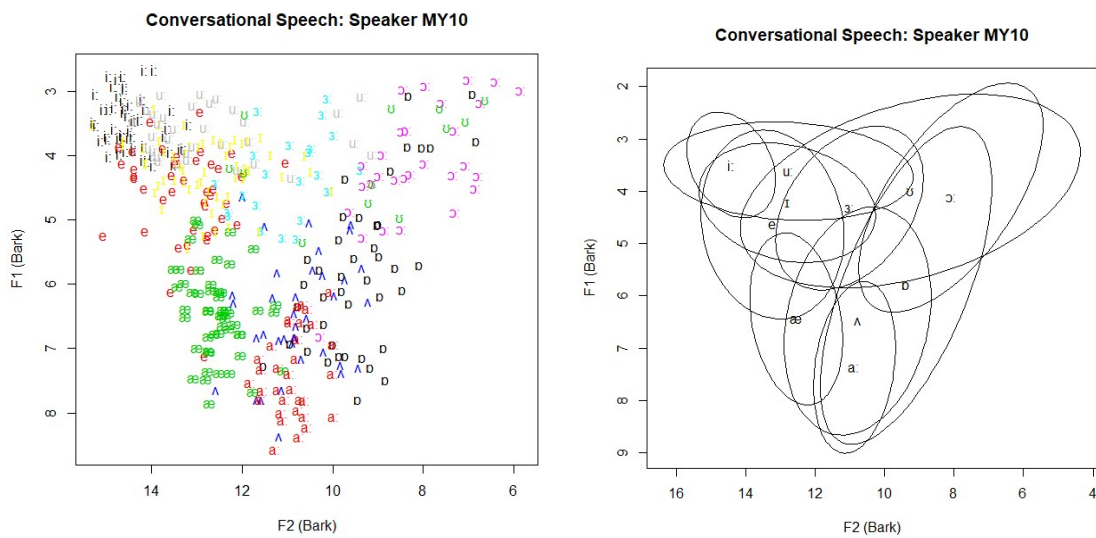

Figure F4. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY06 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.

Figure F5. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY08 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.

Figure F6. Ellipse plot showing vowel tokens (left) and centroids (right) of female speaker MY10 from Mount Roskill. The boundary of each ellipse (right) indicates a 95% confidence interval.
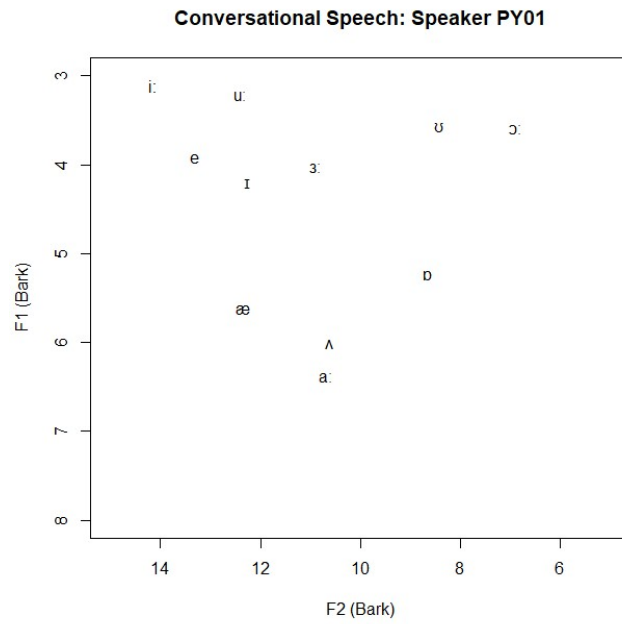
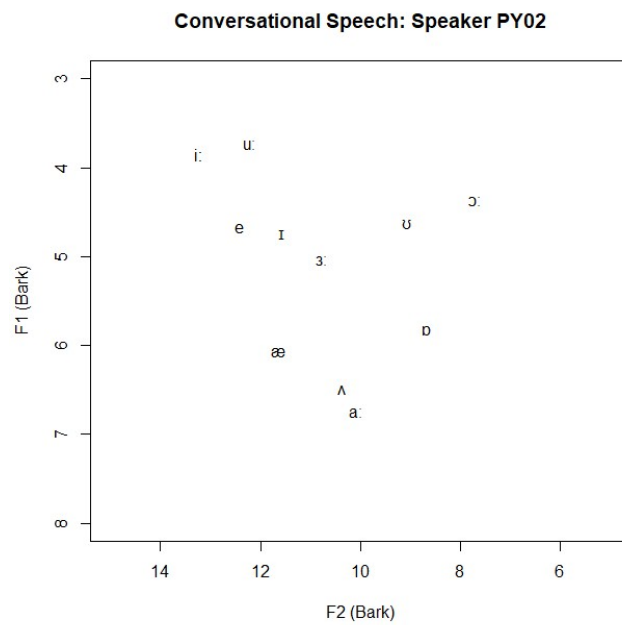*Figure F7.* Papatoetoe young male speaker PY01: F1/F2 space of vowel centroids from conversational speech.

Figure F8. Papatoetoe young male speaker PY02: F1/F2 space of vowel centroids from conversational speech.
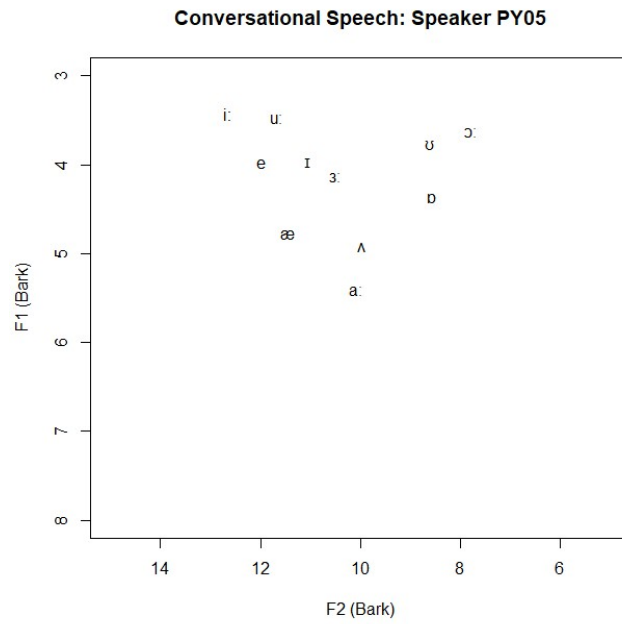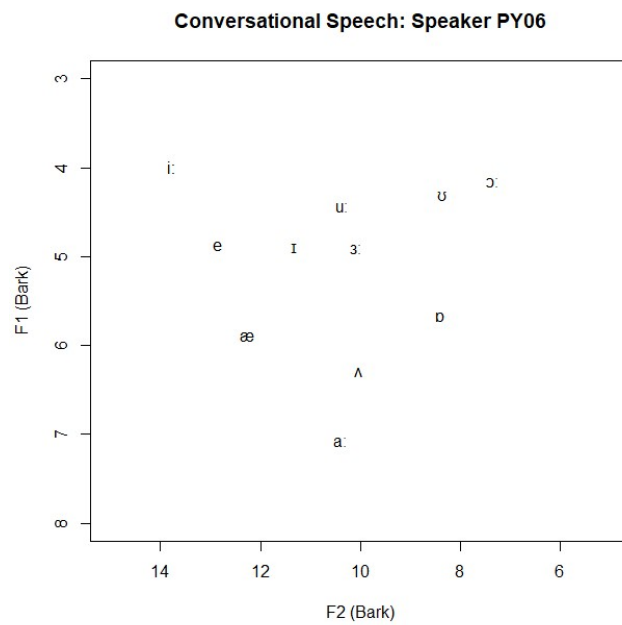
Conversational Speech: Speaker PY05

Figure F9. Papatoetoe young male speaker PY05: F1/F2 space of vowel centroids from conversational speech.

Conversational Speech: Speaker PY06

Figure F10. Papatoetoe young male speaker PY06: F1/F2 space of vowel centroids from conversational speech.

152

Conversational Speech: Speaker PY07

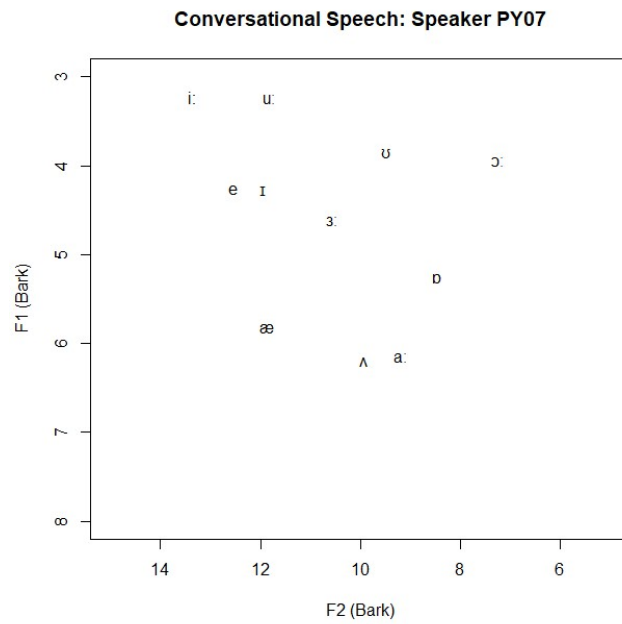*Figure F11.* Papatoetoe young male speaker PY07: F1/F2 space of vowel centroids from conversational speech.

Conversational Speech: Speaker PY10

Figure F12. Papatoetoe young male speaker PY10: F1/F2 space of vowel centroids from conversational speech.

153

Conversational Speech: Speaker PY13

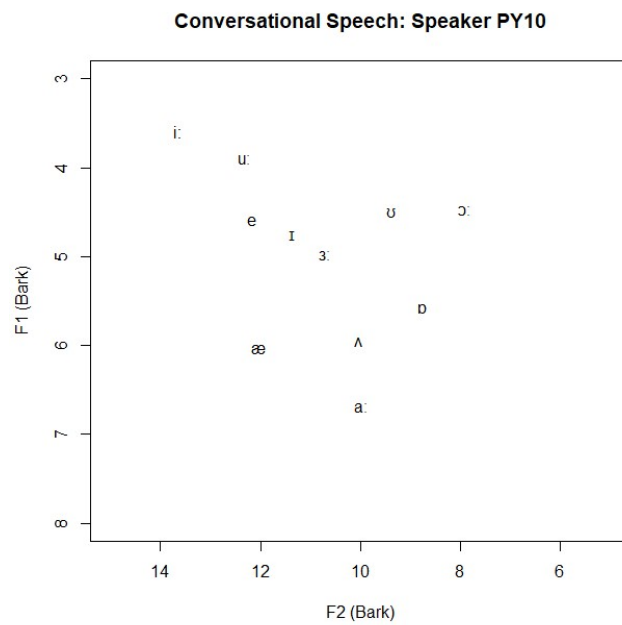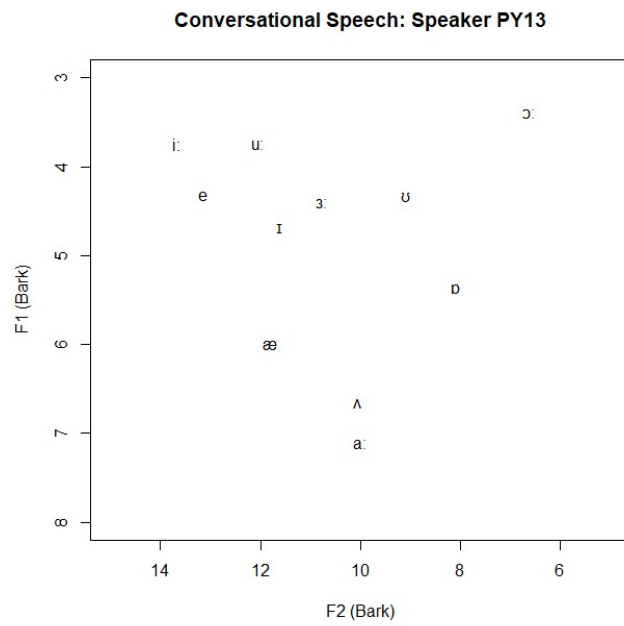*Figure F13.* Papatoetoe young male speaker PY13: F1/F2 space of vowel centroids from conversational speech.

Appendix G

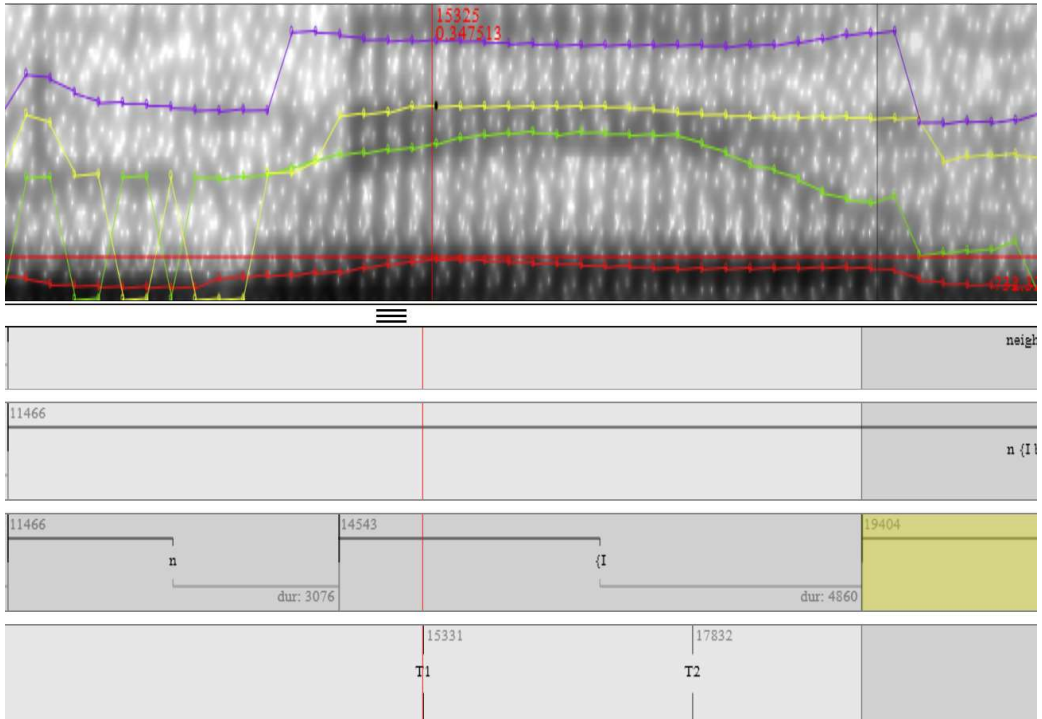Speaker MY03: FACE variation

Figure G1. Speaker MY03: Example of rising FACE vowel. First formant value (red) lowers over time.

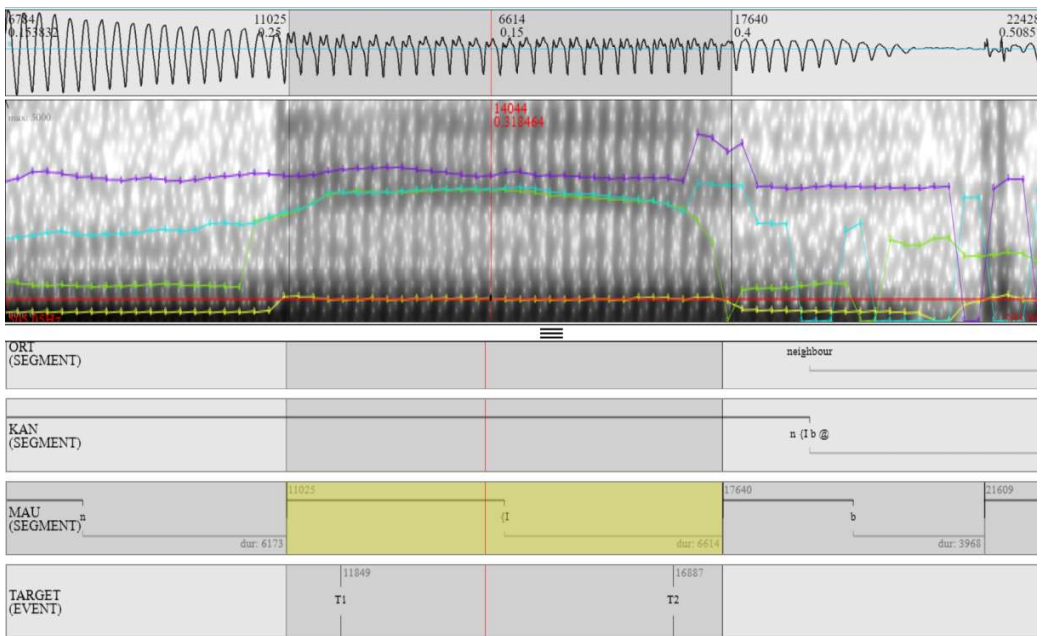*Figure G2.* Speaker MY03: Example of monophthongized FACE vowel. First formant value (red) remains stable over time.

Figure G3. Speaker MY03: Example of falling FACE vowel. First formant value (red) rises over time.

Appendix H

Vowel Onglide: All monophthongs

Table H1.

Vowel Onglide by suburb and gender – Short monophthongs

	DRESS		TRAP		KIT		STRUT		LOT	
	Mean	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**
TOF	0.49	0.14	0.45	0.15	0.48	0.17	0.53	0.11	0.49	0.17
TYF	0.50	0.14	0.46	0.12	0.52	0.18	0.48	0.07	0.46	0.09
TYM	0.51	0.12	0.43	0.14	0.48	0.15	0.51	0.09	0.45	0.09
PYF	0.50	0.15	0.46	0.15	0.52	0.16	0.49	0.09	0.49	0.14
PYM	0.51	0.13	0.43	0.14	0.55	0.15	0.49	0.07	0.45	0.11
MYF	0.52	0.13	0.44	0.15	0.54	0.15	0.45	0.12	0.49	0.14
MYM	0.52	0.14	0.45	0.13	0.48	0.19	0.51	0.11	0.49	0.11
MYF - C	0.43	0.15	0.43	0.16	0.46	0.15	0.46	0.12	0.45	0.16
PYM - C	0.49	0.14	0.49	0.14	0.50	0.13	0.48	0.13	0.44	0.14

Note: Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female(PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female, Papatoetoe Younger Male – Conversational (PYM – C) & Mount Roskill Young Female – Conversational (MYF – C)

Table H2.

Vowel Onglide by suburb and gender – Long monophthongs

	FLEECE		START		THOUGHT		NURSE		GOOSE	
	Mean	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**	**Mean**	**Sdev**
TOF	0.63	0.17	0.45	0.13	0.46	0.14	0.49	0.17	0.43	0.09
TYF	0.56	0.15	0.47	0.09	0.44	0.06	0.44	0.08	0.39	0.11
TYM	0.62	0.19	0.41	0.11	0.50	0.17	0.49	0.11	0.40	0.19
PYF	0.60	0.15	0.49	0.14	0.41	0.20	0.42	0.10	0.41	0.06
PYM	0.56	0.15	0.46	0.15	0.43	0.14	0.49	0.11	0.31	0.15
MYF	0.53	0.14	0.46	0.10	0.41	0.13	0.41	0.15	0.43	0.08
MYM	0.55	0.15	0.50	0.10	0.44	0.12	0.46	0.11	0.43	0.18
MYF - C	0.48	0.15	0.45	0.14	0.37	0.16	0.42	0.14	0.41	0.18
PYM - C	0.56	0.16	0.48	0.12	0.40	0.16	0.50	0.13	0.43	0.14

Note: Titirangi Older Female (TOF), Titirangi Younger Female (TYF), Titirangi Younger Male (TYM), Papatoetoe Younger Female(PYF), Papatoetoe Younger Male (PYM), Mount Roskill Younger Female (MYF), Mount Roskill Younger Female, Papatoetoe Younger Male – Conversational (PYM – C) & Mount Roskill Young Female – Conversational (MYF – C)