Exploring Tasmania's place in Australian English regional variation through diphthongs

Bachelor of Arts, Honours in Language Studies

The Australian National University

Thomas Powell-Davies

October 2022

This thesis is submitted in partial fulfilment of the requirements of the degree of Honours in Language Studies in the College of Arts and Social Sciences.

Declaration

I hereby declare that, except where it is otherwise acknowledged in the text, this thesis represents my own original work. All versions of the submitted thesis (regardless of submission type) are identical.

Chapter 1 partially draws on material submitted for Assignment 2 in LING4106. Chapter 3 partially draws on material submitted for Assignment 3 in LING4106. Chapters 1, 2 and 3 partially draw on material submitted for Assignment 4 in LING4106. Chapters 2 and 3 partially draw on material submitted for the Research Proposal assignment in LING4009. Chapters 2, 3, 4 and 5 partially draw on material submitted for the Research Paper and Research Paper Presentation assignments for LING4009.

Table of Contents

List of Tables

Table of Figures

Acknowledgements

I would like to acknowledge all the time, knowledge and tireless support of my supervisor Dr Rosey Billington over the course of this year. I would also like to acknowledge Dr Ksenia Gnevsheva for her methodological advice through LING4009 coursework, and Prof Catherine Travis for insightful discussions about sociolinguistic variation in Australian English.

Abstract

In recent decades, there has been an increasing focus on regional variation in how English is spoken within Australia. After a long period of research focusing primarily on sociological factors, the focus has broadened to include regional differences, but predominantly based studies within major cities. Tasmania, an island with a distinct sense of identity, can offer valuable insights into the nature of geographically-conditioned linguistic variation in regional Australia. The production of the AusTalk corpus (Estival et al., 2014), a trove of recordings from around Australia, opens the door for this kind of comparison to include some previously neglected places. Diphthongs serve as an interesting point of comparison, as they are often regarded in the public eye as a defining feature of Australian English. This thesis aims to analyse the diphthongs of English as spoken in Tasmania through acoustic phonetic analysis of formant frequencies of diphthongs recorded in hVd wordlist format using the EMU Speech Database Management System (Winkelmann et al., 2017) and R (R Core Team, 2021), and then compare this to more well-studied major city of Sydney. Subsequently, an investigation is performed as to the extent to which there exists location-based variation in the pronunciation of diphthongs between the two cities. This will give a valuable insight not only into Tasmania, but also the understudied areas outside of Australia's major cities more generally.

1 Introduction

This thesis profiles the diphthongs of English as spoken in Tasmania, in comparison to the corresponding data from Sydney, which is the site of the most extensive sociophonetic research on Australian English. The aim is to investigate the extent, if any, to which there are regional differences in diphthong characteristics between the two locations. This is achieved by measuring the duration, and first and second formant frequencies of the vowels, comparing the mean formant values at various intervals over the duration of the vowels, and using linear mixed-effects modelling to determine any statistically significant differences in the data that can be attributed to location. This is hoped to add not only to what is known about English as spoken in Tasmania, but also to the picture of how Tasmania and other areas of regional Australia fit into the broader picture of variation in Australian English.

To date, a large proportion of phonetic research on Australian English has focused on major cities, especially Sydney (e.g. Cox, 2006), but also a select few other major centres including Melbourne (Loakes et al., 2017). Large-scale corpus projects such as the Sydney Social Dialect Survey (Horvath, 1985) and the Sydney Speaks project (Travis et al., In Progress) have broadened the scope of research within the Sydney area. These projects and other more specific work on Sydney have greatly expanded our knowledge of linguistic variation, including phonetic variation, and the nature of socially conditioned linguistic change. In early work on

Australian English pronunciation, Mitchell and Delbridge remarked that "Australian speech is remarkable for its comparative uniformity" (1965, p. 11), after not finding compelling evidence to believe that there was systematic location-based phonetic variation within Australian English. Until relatively recently, the topic of regional variation in pronunciation remained relatively unexplored, when compared with other aspects of variation in Australian English. However, this is no longer the case, with more attention being paid to regionally conditioned phonetic variation in recent years.

This renewed interest has so far resulted in only a limited body of work that focuses on areas outside of the major metropolitan centres, although this too is slowly changing. Tasmania is one area that remains particularly understudied as a state, with very little known about the pronunciation of English as spoken on the island, and it patterns in relation to the rest of Australia. There is still a wide range of questions to be asked about English as spoken in Tasmania, such as how it fits into what is already known about pronunciation in Australian English and to what extent it is aligning with the sound changes that are ongoing in various parts of mainland Australia.

Tasmania presents itself as an interesting region for the purposes of sociolinguistic analysis, as it is a separated island state with a strong sense of identity and place. As language is often used as a way to signify a speaker's membership within groups, it might be expected that there could be ways that Tasmanian identity is expressed through pronunciation differences, as has been observed in various other geographically disconnected regions. This has been an object of study right from the beginning of variationist sociolinguistics, for example in Martha's Vineyard off the coast of Massachusetts, USA (Labov, 1963). However, language use in Tasmania is likely to be shaped dramatically by outside forces, especially given the strong culture that exists of Tasmanians moving to and from the mainland, for a variety of economic, educational and cultural reasons (Easthope & Gabriel, 2008).

There have been coordinated attempts to increase the amount of linguistic data available for a wide range of locations throughout Australia, including Tasmania, in order to facilitate country-wide analyses. This includes the AusTalk corpus (Estival et al., 2014), which is the result of a large project that recorded data with 1000 speakers from 14 locations across every state and both the major territories of Australia. It collected both spontaneous speech data such as interviews and map task recordings, as well as read data such as wordlist and read passages. This corpus is particularly notable in terms of its geographic spread, containing data from both regional and remote Australia. It is also the first corpus publicly accessible on the internet that contains a large enough amount of data from Tasmania to perform a meaningful region-based analysis including the state. The corpus' geographical reach of data collected in a standardised way can enable a wide range of inter-regional comparisons. However, there have so far been relatively few such comparisons.

The aim to profile the diphthongs of Tasmanian English arises from a number of motivations. One of these is to add to the understanding of regional variation within Australian English. Adding information about pronunciation of diphthongs in Tasmania will add another piece of the puzzle of how this variation operates, complementing the work that has already been done by Stanley (2016), documenting the monophthong vowel space of a sample of Tasmanian speakers, also drawn from the AusTalk corpus.

Another motivation for this study is to add to the understanding of the link between language and identity, and how a strong group identity may or may not be reflected linguistically. Tasmania is an example of an island that has a strong identity of its own, but despite this, there is no categorical phonetic difference that Tasmanians are consciously aware of. There is occasionally some public discourse around a Tasmanian regiolect, including an article appearing in the Australian Broadcasting Corporation website (Dalla Fontana, 2019), except this tends to be vague, focusing on general perceptions rather than specific distinguishing phonetic features. This also leaves open the possibility that there is no salient region-based variation in Tasmanian English diphthongs, which would be an interesting finding in its own right.

Having said this, another motivating factor is the only phonetic comparison between Hobart speakers and their mainland counterparts found a noticeable difference in the realisation of monophthongs (Stanley, 2016). It is not clear if diphthong-related variation would be necessarily coupled with monophthong variation in this specific case, so this opportunity remains very much open for exploration. Therefore, the investigation of the realisation of diphthongs is a natural extension of this line of enquiry, helping to paint a fuller picture of the nature, if any, of phonetic variation between Tasmania and the mainland. By expanding what is known about English as spoken in Tasmania, a more accurate picture can be gained about how Tasmanian identity and social networks interact with linguistic factors.

This thesis will achieve its aims by analysing the realisation of diphthongs in a controlled hVd environment. The sample for the study contains 394 tokens from 27 speakers in the age group 18-35, split evenly by gender and location (between Hobart and Sydney). This sample is drawn from the AusTalk corpus project. The sample will be made up of speakers whose first language is English and who completed their primary and secondary education in Australia. The criteria for determining whether a speaker is a member of a location group is that they must have listed their birth location as being within the metropolitan area of the city in which they were recorded.

Throughout this thesis, individual vowels will be referred to using items from the Wells lexical set (Wells, 1982). This set of words is widely used in English dialectology and sociophonetics to stand in place of individual vowels, as example words provide a more neutral reference point than IPA representations that impose a particular standard that may not phonetically align with the variety being described. Conversions between the Wells lexical set and IPA transcriptions of Australian English are available in

Table 2.2.1. The study will focus on the five generally accepted non-marginal diphthongs of Australian English, those being FACE, PRICE, GOAT, MOUTH and CHOICE, for reasons discussed further in Chapter 3.3.

The study is using controlled hVd wordlist data to enable compatibility with as much of the previous work that has been done on Australian English phonetics as possible, in which the hVd frame is the most commonly used environment. It will also provide consistency in terms of many prosodic, intonational and phonetic environment factors, removing some confounding factors that may decrease the chance of having enough data to arrive at a statistically meaningful result working with the limited data that is available. One important such factor is the relative frequency of each vowel, as CHOICE in particular has a relatively low frequency that may result in insufficient data for a meaningful analysis from the available quantity of natural speech data. Once a baseline has been established from this highly controlled wordlist data, natural speech data and all its complexities will be able to be incorporated into future variationist phonetic work involving diphthongs of Tasmanian speakers.

It is outside the scope of this study to use data from a range of locations within Tasmania. There is traditionally a three-way regional split within Tasmania, into the Southern Tasmania, including the urban area of Hobart, Northern Tasmania, including the urban area of Launceston, and North West Tasmania, including the areas of Devonport and Burnie. There may also be differences between the four cities of Tasmania and rural areas. However, there is insufficient data in the AusTalk corpus to be able to investigate this. For this reason, this study only uses data from Hobart speakers, for which enough data is available, and this can be used as a starting point for further inquiry into within-Tasmania regional phonetic variation.

To these ends, the research question to be investigated is: to what extent is there variation in the way diphthongs are pronounced in Hobart, compared to in the more studied mainland location of Sydney?

The literature review in Chapter 2 covers a wide range of topics that background this study and from which this is building from. These include a background of the area of sociophonetics, the properties of diphthongs and vowels more broadly alongside why they are interesting in the context of researching sociophonetic variation, an overview of phonetic research that has been done on Australian English vowels in the past (much of which has focused on Sydney), a summary of what kind of research has been done on Australian English that goes outside the Sydney area, and an explanation as to why Tasmania offers a unique research opportunity, with an outline of the linguistic research that has previously included the island state. Chapter 3 outlines the method used in this study, including information about the corpus that the data is drawn from, the participants and the criteria for determining the sample from which they were drawn, the data itself, how it was processed to prepare it for statistical analysis, and the statistical analysis itself. Chapter 4 contains the results, including tables and plots for the formant and duration values, a brief interpretation of the results for each vowel, and then an explanation of the model and model pruning process that determined which factors are significant in this dataset. Chapter 5 is the discussions and conclusions, unpacking and interpreting the results, making comparisons with previous literature and reflecting on what future work could be undertaken to further investigate this topic.

2 Literature Review

2.1 Sociophonetics and diphthongs

The field of sociophonetics covers the exploration of many topics at the intersection of social identity and the production and perception of speech sounds. Foulkes and Docherty (2006, p. 411) interprets sociophonetic variation as being the way in which "alternative [phonetic or phonological] forms correlate with social factors". Many social factors have been examined, including factors such as age, gender and social class, as well as various regional and location-based factors among many others. In recent times, there have been an increasing number of studies concerned with how social variation and phonetics interact and how an understanding of each of them may allow great progress in the other field (Hay & Drager, 2007, p. 90). In the past, the equipment needed to conduct acoustic analysis was difficult to obtain, but rapid technological advances over the last several decades mean this is no longer the case, so there has been an increase in studies that can quantify fine phonetic details and their relations to social variables (Hay & Drager, 2007, pp. 91–92). There are many different ways variation in linguistic production can be captured quantitatively. For vowels, as well as other variables such as duration, the most studied variables are formants, which are the various resonant frequencies manipulated by movement of the vocal tract. The first (F1) and second (F2) formants above the base pitch of the voice (F0) are the most often examined variables for determining vowel quality, although the other formants may also offer many different insights, depending on the nature of the variation that is being examined. F1 approximates an indicator of how high the tongue is within the mouth, and F2 approximates an indicator of how far forward it is (Ladefoged & Johnson, 2014, p. 206).

One of the foundational researchers at the intersection of sociolinguistics and phonetics was William Labov, notably including his study on vowels produced by English speakers in Martha's Vineyard (Labov, 1963). This work was a key development in the quantitative study of the interaction between social factors and phonetic variables and much of the methodology developed is still in use today. The study collected data from 69 participants, representing a wide demographic range within the island population, by occupation, ethnicity and location within the island. The PRICE and MOUTH vowels were undergoing a shift on the island, in which the initial target of the diphthong was being centred. Through his analysis, Labov demonstrated that the first and second formants of this initial target correlated with a number of factors, such as age, occupation, location and attitudes to island life. By using statistical methods to demonstrate a relation between social factors and the degree to which residents of Martha's Vineyard adopted this local shift in diphthong pronunciation, he was able to show how a sound change can be observed and measured scientifically and synchronically.

In sociophonetic research, there has been some research on consonants, prosody and other features since then, but vowels have been a major largest focus. However, much of this research has focused on monophthongs, typically measured by taking formant values at the midpoint or calculating a vowel target. There has been less sociophonetic research focusing on diphthongs. This is because, for earlier studies, they presented a greater challenge, due to the analytical difficulties in analysing their highly dynamic trajectories. Diphthongs have previously been defined as vowels that "have two separate targets" (Ladefoged & Maddieson, 1996, p. 321) as opposed to monophthongs that are traditionally thought to only have one. This means that, when articulating a diphthong, the tongue moves from one position to another over the course of the vowel. Within diphthongs, there are a number of distinctions that can be made. One major distinction that is relevant to Australian English is based on the general direction the diphthong trajectory moves in in terms of vowel height. There are the closing diphthongs such as FACE, PRICE, GOAT or CHOICE, opening diphthongs such as MOUTH, and centring diphthongs, such as the marginal diphthongs NEAR and CURE. Another important concept in the classification of vowels is vowel inherent spectral change, or VISC. The concept of VISC captures the fact that all vowels, including monophthongs, have some degree of movement inherent to them. This concept has received some attention (e.g. Morrison & Assman, 2013), including in the context of Australian English (Watson & Harrington, 1999). The distinction of vowels between monophthongs and diphthongs is therefore not so clear cut, with some works such Elvin et al. (2016) choosing to refer to them as "nominal" diphthongs and monophthongs instead. For clarity, this study will not include the word nominal at every mention of diphthongs or monophthongs, however recognising that these categories are not concrete and fixed groupings. The way in which the selection of diphthongs was made in the context of this study will be discussed further in chapter 3.3.

2.2 The vowels of Australian English

Sociolinguistic work on the description of vowels in English as spoken in Australia goes back to the pioneering work of Mitchell and Delbridge (Mitchell, 1946; Mitchell & Delbridge, 1965). Mitchell's initial work (1946) described the pronunciation features of Australian English, treating it as a separate variety. This work was built on significantly in collaboration with Delbridge for the second edition (1965), which went into detail about social variation alongside its general description of Australian English. While detailing the development of English in Australia, they were one of the first to recognise that Australian English was its own variety that is spoken in the community, and not a subvariety of the English as spoken in southern England (1965, p. 59). The corpus used for this book was drawn from tape-recorded data with 7,736 students in their final year of school, collected from every state in Australia. Included in this sample were 289 students from Tasmania (Vonwiller et al., 1998). The aim of their research was to arrive at a classification for different types of speech found within Australia. As well as the linguistic data, they also collected a wide variety of social data in order to search for correlations. From their observations, they came to the conclusion that there was no detectable systematic regional variation in their data on Australian English (1965, p. 11).

However, they arrived at a three-way classification that primarily reflected socioeconomic background, with the categories named as broad, general and cultivated (1965, pp. 14–15). The six vowels used in their classification to distinguish between the accent types, described varyingly as monophthongal or diphthongal depending on the sociolect, are the vowels of the words FACE, GOAT, PRICE, MOUTH, FLEECE and GOOSE (1965, pp. 40–44). This shows that, even from the early analyses, diphthongs have been very important in indexing social variation in Australian English. This classification of Australian English vowels, alongside the more modern HCE system (Harrington et al., 1997) and their Wells lexical set correspondences (Wells, 1982) are included in

Table 2.2.1.

Table 2.2.1 – Vowel IPA symbols in the Mitchell and Delbridge system and the Harrington, Cox and Evans system alongside corresponding items in the Wells lexical set

Mitchell & Delbridge system	HCE system	Wells lexical set
/i/	/iː/	FLEECE
/ɪ/	/ɪ/	KIT
/ɛ/	/e/	DRESS
/æ/	/æ/	TRAP
/a/	/ɐː/	BATH or PALM
/ɒ/	/ɔ/	LOT
/ɔ/	/oː/	THOUGHT
/ʊ/	/ʊ/	FOOT
/u/	/ʉː/	GOOSE
/ʌ/	/ɐ/	STRUT
/ɜ/	/ɜː/	NURSE
/ə/	/ə/	LETTER or COMMA
/eɪ/	/æɪ/	FACE
/oʊ/	/əʉ/	GOAT
/aɪ/	/ɑe/	PRICE
/aʊ/	/æɔ/	MOUTH
/ɔɪ/	/oɪ/	CHOICE
/ɪə/	/ɪə/	NEAR
/ɛə/	/eː/	SQUARE
/ʊə/	/ʊə/	CURE

Subsequent studies started using more quantitative methods instead. A pioneering acoustic study of Australian English by Bernard (1970) looked closely at a sample of 171 adult male speakers that had been classified in roughly equal proportions into the Mitchell-Delbridge three-way structure. The speakers were all from New South Wales, mostly the Greater Sydney area, and the sample contained high school and university students as well as people with other occupations. The formant frequencies of both diphthongs and monophthongs (excluding schwa) in the environment hVd were detailed at three points along each vowel and noted along with the duration, allowing a statistically based description and internal comparison of Australian English vowels. This comparison also included an analysis of diphthong targets in relation to the monophthong vowel space. This backed up Mitchell and Delbridge's (1965) theories about diphthong targets varying by sociolect, specifically analysing the vowels of FACE, PRICE, MOUTH and GOAT. A reanalysis of the Bernard data by Cox (1998) was performed, using modern apparent-time methods to pay special attention to the interaction with age and the formant values, in order to identify any sound changes that may have been in progress at the time of Bernard's data collection in the late 1960s, flagging potential changes in progress for the vowels of GOAT, GOOSE and NEAR.

In more recent decades, impressions of a changing linguistic landscape prompted a push to update the description of Australian English vowels, as well as an effort to start describing more of the variation that occurs within Australian English. The study produced by Harrington, Cox and Evans (1997) filled this gap and still forms the basis of the current standard for Australian English IPA transcriptions. The data is drawn from the Australian National Database of Spoken Language (ANDOSL) (Vonwiller et al., 1995). This heavily-used corpus has a selection of participants that have been balanced for age, gender and Mitchell-Delbridge accent type, but the speakers were all from the Sydney area. In total, the sample drew 2230 tokens from 119 speakers. One interesting finding from this was that many of the indicators that were previously used to determine whether an accent fits into broad, general or cultivated were no longer reliable measures. For example, for the male speakers, there were no statistically significant differences found for the second target of any diphthongs according to the Mitchell and Delbridge accent classification of the speakers. They also found that the Mitchell and Delbridge IPA transcriptions were no longer representative of their current Australian English sample, and suggested replacement transcription conventions for a number of phonemes, as previously listed in

Table 2.2.1.

Another study was conducted by Cox and Palethorpe (2001). They took a subset of data from Sydney male school students collected by Bernard in the 1960s and compared it with a demographically similar sample from the 1990s, again drawing on a sample of male school students from various areas of Sydney. A statistically significant difference between the older Bernard data and the newer data was observed for 14 different targets, with four diphthongs represented: FACE, PRICE, MOUTH and GOAT (Cox & Palethorpe, 2001, p. 24), confirming clear evolution in the vowel system of Australian English over the decades. In particular, the diphthongs of FACE and GOAT play a role in the ongoing shifts in Australian English, linking the ongoing change in GOAT to the ongoing GOOSE fronting that had been observed in multiple locations. Their study acknowledges that there may be regional effects between different areas of Sydney at play in the corpora, but as the paper only focuses on Sydney, so does their region-based analysis. The data was also re-examined later to include a detailed description of vowel length and even the length of each section of the diphthong trajectories (Cox, 2006b). They found that males have statistically significantly longer onglides especially for tense vowels and diphthongs, and that duration provides a point of difference between the vowel pairs of KIT and NEAR as well as DRESS and SQUARE, which do not otherwise appear to be differentiated by formant values. Other works by Cox and colleagues also focused on Sydney data in depth, including a general description of the features of Standard Australian English (Cox & Palethorpe, 2007) and a textbook on Australian English Pronunciation and Transcription that relies on formant data from vowels extracted in the hVd environment from 34 young adult speakers from Sydney (Cox & Fletcher, 2017).

Studies since then have moved their focus away from a strong three-way distinction in accent type, as, in the context of current Australian English speech, it is no longer as salient of a way to describe variation, due to a widespread convergence to an accent style more consistent with what might have previously been described as General Australian English (Cox, 2006; Price, 2008). Although they remain lesser studied, several of these have included work on diphthongs. One important study took data from 19 young adult speakers in Western Sydney and looked at the impact various different environments had on formants and duration (Elvin et al., 2016). As previously mentioned, this study importantly made specified vowels as only nominally being monophthongs or diphthongs, in recognition of the fact that all vowels have some sort of dynamics. Alongside some other studies, it grouped NEAR with the nominal monophthongs, as it patterns more similarly to them. On the other hand, the FLEECE vowel had by far the largest onglide of any vowel (Elvin et al., 2016, p. 177). Due to trajectories like this being observed in data collected from Sydney, the FLEECE vowel is sometimes treated as a diphthong. This study also showed tangible impacts that the /hVd/ context has on the properties of vowels, including diphthongs, especially with regard to lengthening the duration of the vowel segment.

Other ongoing projects have also made substantial contributions to understanding of Australian English sociophonetics, including of diphthongs. Perhaps the most significant currently developing corpus is the Sydney Speaks project (Travis et al., In Progress). This corpus takes data from all over Sydney with participants from a variety of social backgrounds, with a particular focus on different ethnic communities. Studies that draw data from this project do so with reference back to the Sydney Social Dialect Survey of the 1970s (Horvath, 1985). One contemporary study on diphthongs that has come out of this project looks at change in some diphthongs of Australian English and recognises that a shift is ongoing, led by middle class and ethnic Chinese Australians (Grama et al., 2021). Interestingly, this study includes the FLEECE vowel as a diphthong, as well as the more traditional diphthongs of FACE, GOAT, MOUTH and PRICE. The data drew from both the Sydney Speaks project and the Sydney Social Dialect Survey, with a total of over 23,000 tokens from 173 speakers, both male and female. The Sydney Speaks data selected for this study include members of the Anglo, Italian and Chinese communities, while the Sydney Social Dialect Survey data only included Anglo and Italian communities. The studies show a general trend away from realisations traditionally ascribed to the broad sociolect over time for all five of the vowels to some extent, but less so for PRICE and MOUTH, with the current Chinese-background speakers being furthest away from the older realisations. This reflects a broader demographic trend within the Chinese community sample collected by the Sydney Speaks project, in which there are minimal Chinese participants from a low socioeconomic background. This study forms part of a greater body of work that indicates diphthongs are an interesting area in which indexing of social factors through an ongoing change is apparent.

2.3 Looking beyond Sydney

Many of the key acoustic phonetic studies on variation in Australian English have largely drawn their data from Sydney. More recently, studies have increasingly drawn data from other parts of the country, widening the scope of this research.

Much of this work has looked specifically at vowels. Nearly all of the variation captured has been phonetic, but there has also been some analysis of regionally conditioned phonemic features. One notable study by Bradley (1991) examines the TRAP-BATH split, or the distinction between what corresponds to /æ/ and /ɐː/ in the HCE transcription system (Harrington et al., 1997), in five Australian cities, capturing phonemic variation in the pronunciation of words such as "chance" and "demand". It is notable that this is one of very few studies that Hobart was included in, with Hobart having the highest rates of pre-nasal TRAP vowel of any of the cities, almost categorically so before nasals, and thus being furthest away from British Received Pronunciation. On the other end of the spectrum, almost all speakers from Adelaide used the BATH vowel in similar environments. There is also an emerging phonological change is also occurring in parts of Victoria, of a merger of DRESS and TRAP vowels before /l/ (Loakes et al., 2017), which correspond to /e/ and /æ/ in the HCE transcription system (Harrington et al., 1997). This has been occurring in parallel with some other varieties of English including New Zealand, and has received some attention, being dubbed the CELERY-SALARY merger. Further investigation into a difference between regional and rural Victoria and New South Wales has been undertaken by Cox and Palethorpe (2004), in which identifiable differences between the Victorian regional centre of Wangaratta and the New South Wales locations of Wagga Wagga, Junee and Temora supported existing theories around, for example, GOOSE fronting being more advanced in New South Wales, and that Wangaratta speakers had a more retracted DRESS vowel before /l/, indicating the beginnings of a CELERY-SALARY type merger. Another study has shown that Warrnambool, in Victoria's southwest, is much more advanced in the CELERY-SALARY merger than Albury-Wodonga, a border city between northeast Victoria and New South Wales, both in terms of perception and production (Loakes et al., 2014). Warrnambool's relatively completed merger is also visible in comparison to Mildura, a northwestern city on the border with New South Wales (Loakes et al., 2018).

Besides investigations of the CELERY-SALARY merger, there has been some more work conducted on Australian English speakers in Melbourne

(Billington, 2011). This study looked at just monophthongs within the hVd context. By comparing modern data from 22 high school aged speakers (13 female and 9 male) in Melbourne and comparing it with other modern data taken from studies conducted in Sydney (Cox, 2006b) and Adelaide (Butcher, 2006) as well as historical data from New South Wales (Bernard, 1970), the study contributed not only to understanding the nature of regional variation in Australian English, but also some limited perspective into evolution over time. An important finding was that different genders oriented to region-based variation in different ways, as well as several New South Wales-specific innovations.

There have been several other comparative phonetic studies of Australian English vowels, looking at various locations around the country. One paper of this nature looks at hVd vowels of 92 female young adult Adelaide speakers and compares them to a similar sample of young adult Sydney speakers from Cox's study (2006b), finding that Adelaide appears to be more conservative in terms of ongoing sound changes, for example GOOSE fronting, and potentially having a more close space of articulation in general (Butcher, 2006). The emergence of this closer articulation was further re-examined in a later paper, where data from 70 young adult female speakers from Adelaide was compared with what had been gathered in three previous studies from different time periods (collected in 1945, 1991 and 2002), totalling 117 additional speakers (Butcher, 2012). It found interesting findings for example in GOOSE fronting, that showed that Adelaide was not as advanced in this sound change as the Sydney reference data, and that the diphthong trajectories that were analysed are shortening over time, except for the GOAT vowel.

One other paper introduces Perth into the mix - Docherty et al. (2019) examine short front vowels of young adult Perth speakers, as obtained using conversational data, and compares the results with what has been found previously in other cities. It does this by drawing on a range of literature that takes data from Melbourne (Billington, 2011), Western Sydney (Elvin et al., 2016) and Sydney more broadly (Cox, 2006b). The means for all three of the examined vowels, those being KIT, DRESS and TRAP, were less fronted than in the other studies, but this is connected to the nature of using natural speech data: the study also demonstrated that duration of the vowel had a significant effect on the degree of reduction. An earlier iteration of the corpus used in this study was also examined in an earlier study that examined the effect of checked syllables on vowel trajectories, achieving this by treating all vowels in the study, including the traditional monophthongs, as dynamic (Docherty et al., 2015).

However, there has been a severe lack of research on anything outside of the major cities. Regional and remote Australia as defined by the Australian Bureau of Statistics (2016) is a very diverse collection of places. In addition to the previously mentioned studies involving Warrnambool, Mildura and Albury-Wodonga in investigations of the CELERY-SALARY merger, one of the studies that does use data from regional Australia is on /s/ retraction in Australian English (Stevens et al., 2019), takes its sample from Braidwood, a town situated approximately 60km east of Canberra, examining how a sound change spreads through a community. Interestingly, it found that exposure alone was an effective predictor of the degree to which an agent picked up the change. Another study on the production of voiceless plosives by primary school aged speakers took its data from a sample of 18 speakers from Yarrawonga, a town in northeast Victoria (Ford & Tabain, 2016).

There is also the AusTalk corpus (Estival et al., 2014) which has been the result of a nationwide coordinated project to produce a standardised set of linguistic data from around the country. It includes sizeable amounts of data, with a total of 1000 speakers between every state and major territory of Australia, making region by region comparison a more accessible avenue of research inquiry. It also includes a good amount of data from regional and remote areas, being the first publicly accessible linguistic corpus to do so. More details about the corpus are provided in Chapter 3.1.

One paper that utilises this resource is by Cox and Palethorpe (2019). This paper compares vowel spaces and trajectories in the hVd environment, using data drawn from the AusTalk corpus. The four most commonly studied cities of Sydney, Melbourne, Adelaide and Perth are included in the study, and statistically significant differences in trajectory are identified for a limited number of vowels. They found varied differences between the cities, which were interestingly also not consistent across genders.

2.4 Tasmania in the picture

Very little of the island state of Tasmania features in any research on English. Despite this, the geographical separation and distinct sense of identity that islands bring providing an interesting setting for geographically conditioned changes to come into being. This has been noted since the very early days of sociophonetics, with Labov's previously mentioned landmark work (1963) taking place on Martha's Vineyard, an island off the coast of Massachusetts, USA. The diphthongs of English as spoken in islands has been highlighted on numerous occasions since then. Many of these studies have focused on smaller islands compared to Tasmania, which means they are likely to have a closer-knit island community and identity, but may also be more reliant on their respective neighbouring regions. The Falkland Islands, a UK territory off the coast of Argentina, is another interesting example of this. The variety of English there exhibits much variation between speakers and even within speakers, but it still maintains many distinct characteristics, including in the pronunciation of vowels (Britain & Sudbury, 2010). Another, albeit quite dated study, investigates English as spoken on Cape Barren Island, an island situated in the Furneaux Group, off the north-east of Tasmania. It identifies a variety of different factors that contribute to a distinct variety from Standard Australian English, not only in terms of the island dynamic, but also some features connected to the Indigenous heritage of most of the island's inhabitants (Sutton, 1975). However, the relatively closed network status of islands has also been critically investigated. For example, on Mersea Island, off the coast of Essex, England, it has been found that a widening of islanders' social networks to include more interaction with the mainland has resulted in realisations closer to what is found in neighbouring mainland areas (Amos, 2011).

In terms of regional identity, Tasmania is very much uniquely positioned within Australia, as a separate and clearly defined island, maintaining a relatively strong in-group versus out-group mentality, despite a large number of Tasmanians spending time in other places for economic or educational reasons. In their examination of the identity of young Tasmanians who have spent time on the mainland, Easthope and Gabriel describe Tasmania as being "both simultaneously bounded and networked" (2008, p. 173), referring to the juxtaposition of island identity within a context of normalised migration.

In the 1990s, there was interest in a system of assigning gender to inanimate objects in Tasmanian English, for example using the pronoun "he" with a turnip as the referent, undertaken by Pawley (e.g. 2002) and Wierzbicka (2002). The discussion between these authors described a system that was still in practice in Tasmania to a greater extent than had been described in other parts of Australia and New Zealand, with Tasmania being described as a linguistically conservative area in this regard, holding onto an Australia and New Zealand-wide innovation for longer than the rest of the places. Wierzbicka also gave a number of reasons for why she believed Tasmania was linguistically conservative, including its island status, but also its decentralised population and lack of international migration-driven population growth. Although not quantitative in nature or connected to phonetics, these papers are important as the earliest, and some of the only, literature available specifically on English as spoken in Tasmania.

Tasmania was included in a comprehensive study by Bryant (1992) on region-based lexical variation in Australian English. In this thesis, based on the way it patterned for various lexical items, Tasmania was generally included with Victoria and the Riverina region of New South Wales as part of the greater south-east Australia region. The study identified three main patterns that Tasmanian usage fell into. These were following the south-east Australian mainland, having its own individual term, and not having a word for an item as it does not exist in Tasmania. The grouping with Victoria is consistent with what was found in a later online survey with over 22,000 respondents, in which Tasmania patterned with Victoria in the majority of cases (Billington et al., 2015). A large, more recent study on folk perceptions of regional variation in Australian English that also drew on these works also included Tasmania as a distinctive region as a result of the interviews that were conducted. Tasmania was labelled as a distinct speech region by only a minority of respondents to the survey, but those who did nominate it assigned it labels that combined a perceived Britishness with nonstandardness, remoteness and uneducatedness (Kingstone, 2019, p. 110).

There is one comparative phonetics study that has been undertaken on Tasmanian English vowels (Stanley, 2016). Drawing a sample of 1,096 tokens from the AusTalk corpus, Stanley analysed the monophthongs of Hobart speakers from two age groups: 24 younger speakers aged 20-39 and 15 older speakers aged over 60 years, of the full total 18 being female and 21 being male. He compared the Hobart speakers with comparable samples from Sydney, taken from Cox (2006b), and Melbourne, taken from Billington (2011). It is of particular interest that a generally closer and backer vowel space was found for nearly all of the vowels when comparing with the major mainland cities. Another unpublished manuscript by Stanley (n.d.) examined the monophthong space of Tasmanian English more closely, also providing formant values for each vowel, with original data recorded from three male speakers from Hobart, aged between 35-37.

This leaves the study of diphthongs as a natural next step for this. As previously mentioned, diphthongs form an interesting point of focus for a large range of sociophonetic variation, yet, despite this, they remain unstudied in the Tasmanian context. An analysis of the formant values of diphthongs as pronounced by Tasmanian speakers and a comparison of this with data from a major mainland centre that has been more comprehensively studied, such as Sydney, contributes tangibly to the development of an understanding of Tasmania's place in Australian English variation.

3 Method

3.1 Corpus

The data used in this study is drawn from the AusTalk corpus (Estival et al., 2014). This arose from a large project, carried out between 2011 and 2014, which aimed to develop a new iteration of corpora such as ANDOSL (Vonwiller et al., 1995) or Mitchell and Delbridge (1965), in that it provides a database of speech materials from a wide range of participants across Australia. The data was collected via a network of 13 participating universities all around the country. Data collection took place in 14 cities in all states and both of the major territories of Australia. These included all the capitals, as well as other regional and remote locations. The data is publicly available on the internet, after registration, through the Alveo portal (Cassidy et al., 2014).

The data collection at each location was in accordance with a standardised recording setup and protocols. The setup consisted of a 'black box', a portable computer with specialised software to display prompts (*About AusTalk*, 2012). Each research assistant that was involved in the data collection participated in a workshop where they learnt how to use the equipment. This standardised setup allows the data to be used for a comparison between locations with minimal concerns about confounding variables. The corpus contains data elicited through a variety of different prompts, including both spontaneous and read speech. This allows for data to be used to examine language in a variety of different contexts to build a clearer picture of the variation that is occurring, as well as applications in other fields, such as speech and language technology. The AusTalk data remains the only digitised and publicly available corpus that contains a useful sized sample of Tasmanian speakers.

3.2 Participants

The sample for this study contains a total of 43 participants. The most appropriate inclusion criteria were determined after browsing the basic demographic information that is available about each speaker on the Alveo portal (Cassidy et al., 2014). The participants in the sample are all in the age range 18-35, in order to reduce any confounding effect of age as much as practical. One of the criteria for participating in the AusTalk study is that the speaker must have completed their primary and secondary education in Australia, in order to reduce any confounding effect of foreign or learner-accented speech (Estival et al., 2014). To approximate who would represent a sample of participants meaningfully connected to their city, Hobart speakers were considered to be those who were recorded at the University of Tasmania and listed their birth location as Hobart. The Sydney sample was drawn from those who were recorded at the University of Sydney and listed their birth location as somewhere within the metropolitan New South Wales area as defined by the NSW Government Department of Planning and Environment (2020). If all those whose residential history indicates they had spent any time living outside these areas were excluded, the sample for the Hobart speakers would become undesirably small. For this reason, the detailed residential history of the participants was not used as an exclusion criteria.

Following these inclusion criteria, the sample for Hobart is narrowed to 13 speakers: 6 female and 7 male. For Sydney, it leaves 22 speakers eligible: 7 female and 15 male. All eligible female participants were initially included in the Sydney sample, but upon inspection of the data it became apparent that one of them did not complete the wordlist task, so the sample needed to be reduced to 6. For the males, 7 speakers were selected with ages that matched the Hobart speakers as closely as possible, with the lower numerical AusTalk speaker ID being used as a tiebreaker in the event of two eligible speakers with the same age. Upon data inspection, it became clear that one of these had also not completed the wordlist task, so this speaker was replaced with another speaker of the same age. An eighth speaker was randomly selected from the remaining speakers. The final distribution of speakers by gender and location is listed in Table 3.2.1.

Table 3.2.1 – Breakdown of speakers by gender and location

	Female	**Male**	TOTAL
Hobart	6	7	**13**
Sydney	6	8	**14**
TOTAL	**12**	**15**	**27**

3.3 Data

The data for this study is drawn from the read speech section of the AusTalk corpus, more specifically the wordlist. In this task, the participants were asked to read out a list of prompts. The wordlist contains over 300 items, that were specifically picked to capture a wide range of patterns, both segmental and prosodic. The wordlist was recorded a total of three times over different sessions, with the wordlist being shuffled each time. The data being used for this study are diphthongs inside an hVd environment, as this has been the most frequently used environment in Australian English phonetic studies over time, so provides comparability with previous work. The prompts used to elicit these were "hade" (FACE, or /æɪ/ in the HCE system for phonemic transcription of Australian English (Harrington et al., 1997)), "hide" (PRICE, /ɑe/), "hode" (GOAT, /əʉ/), "howd" (MOUTH, /æɔ/) and "hoyd" (CHOICE, /ɔɪ/).

These hVd words contain the vowels that have historically been represented as diphthongs, with the notable exclusion of the CURE vowel, which was not present in the corpus in a hVd frame, and the NEAR vowel, which was. There is some discussion as to the status of the NEAR vowel being a diphthong in Australian English, and, for many speakers, it is quite monophthongal. NEAR produced in the hVd context spoken by Tasmanian speakers in the AusTalk corpus has already been examined in Stanley's (2016) paper on monophthongs, and was shown to have quite a short trajectory. Additionally, in the present study, there were a high number of problematic tokens from the prompt "heared", likely due to this being an ungrammatical word for many speakers. Many participants misread or misunderstood the prompt and produced a token with the NURSE vowel instead, as in the more standard word "heard". There were also a smaller number of tokens when the participant used the SQUARE vowel. Also, some speakers pronounced this prompt with a rhotic following the vowel, also rendering the tokens too segmentally different from the intended target environment. For these reasons, the tokens collected using the prompt “heared” were not included in this study.

Several of the tokens were missing, perhaps as the participant may not have completed all the recording sessions. A number of additional exclusions also had to be made, as there were also three tokens where the word recorded did not match with the prompt, one file that was completely empty and, in one case, the audio file had been corrupted to the point of unintelligibility. After all of these exclusions, the total token count was 394.

3.4 Data processing

The data was downloaded from the portal Alveo (Cassidy et al., 2014), where it comes with 7 different tracks. One of these is in the format of 32-bit 44.1kHz stereo audio, which was converted to the standard 16-bit 44.1kHz mono format typically used in speech research. This is done using the software Praat (Boersma & Weenink, 2022). For some of the files, an error message was generated that some samples had been clipped from the file. However, this only happened for 8 of the tokens, and in every case the total proportion of samples clipped was below 0.2%, often significantly lower. This error was therefore not considered a concern and the tokens were still included.

After the data was downloaded and converted to the standard format, a script (Crosswhite et al., 2013) was used to assist in generating text grids to correspond to each file, and then the boundaries of the segments were manually annotated. The beginning of the /h/ and the end of the /d/ is largely irrelevant for the purposes of this study, but the initial boundary of the /h/ was annotated at the earliest point that there was visible greater disturbance than background noise in the spectrogram, and the final boundary of the /d/ was annotated after the release and initial burst of air after it, if there was a release present. The end of the /h/ which is the beginning of the vowel is annotated at the crest of the first regular waveform after the aperiodic noise of the preceding fricative. The end of the vowel which is the beginning of the /d/ is set at the crest after last waveform that is recognisable as being part of the regular high-amplitude periodicity of the vowel, as opposed to the visible closure that corresponds with the /d/.

The next stage of the process was undertaken using the EMU Speech Database Management System via the emuR package (Winkelmann et al., 2017) for the R language (R Core Team, 2021), which, for this study, was coded, compiled and executed in the RStudio environment (RStudio Team, 2021). The word and phoneme segment tiers from the Praat text grids were converted to an EMU-DB hierarchical database together with the .wav files, and the tiers of the linked up in a one-to-many relationship. An R script (Winkelmann, 2017) that takes the .wav files and uses the formant tracking algorithms from Praat was then used to generate formant tracks that are compatible with the EMU-DB software. This was algorithm was found to be greatly better suited to the data than EMU-DB's inbuilt formant tracking function From these formant tracks, F1 and F2 measurements were extracted at 10 points along the vowel, at every 10% from 5% until 95%. Any measurements of 0Hz were assumed to be formant tracking errors and removed from the dataset.

3.5 Analysis

Following a preliminary look at the general trajectories of the vowels, based on the raw formant data from the Hobart and Sydney speakers, 15% into the vowel was selected as the most useful measurement point for the initial portion of the vowel, as measurements at the 5% interval would likely still be under the influence of frication from the /h/. Towards the end of the trajectory, 85% into the vowel was selected as the most useful measurement point for the final portion of the vowel, to minimise influence by the following /d/.

The initial measurement point, final measurement point and duration are all important measures in determining the properties of each diphthong. These form the dependent variables for analysis in this study. The independent variables include location, whose effect is the object of study as outlined in the research question, and gender, which is already known to have a very large effect on speech patterns. Due to potential social phenomena that may have different norms for different genders, it is also important to consider the interaction between location and gender in any models produced. An additional independent variable, speaker, was included as a random effect. Therefore, for the purposes of this study, a linear mixed-effects model is the most suitable for analysing difference in formant values.

As is standard, $p < 0.05$ will be used as the value to determine statistical significance. With several models being analysed, there is a fairly high chance that statistical hypothesis errors will be committed. If location had no effect on diphthong formant values, there would be a reasonable chance that a Type I error would be committed, that is, incorrectly rejecting the idea that there is no effect. There is also the chance for Type II errors to be committed, that is, assuming that there is no effect when there actually is one. One factor that may lead to this is the sample size: it is possible that a model using the number of tokens that are included in the present study are not powerful enough to detect a given effect.

4 Results

4.1 FACE vowel

The first vowel to be analysed is the FACE vowel, which corresponds to /æɪ/ in the HCE system for transcribing Australian English (Harrington et al., 1997). These tokens were elicited using the prompt "hade". There are a total of 80 tokens in the sample for this vowel: 18 from Sydney females, 24 from Sydney males, 18 from Hobart females and 20 from Hobart males. Figure 4.1.1 illustrates the mean F1 and F2 trajectories of the vowel separated by location and gender, with Hobart speakers in blue and Sydney speakers in red.

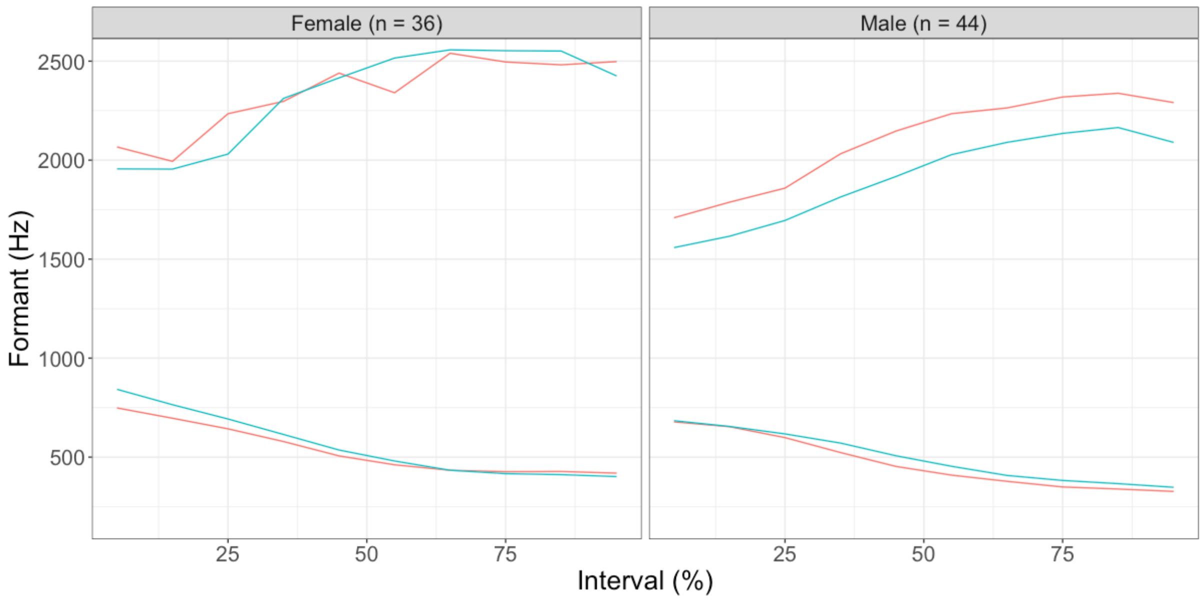

Figure 4.1.1 – Mean F1 and F2 trajectories for FACE, separated by gender and location

There is no large visible difference in formant values between location for the female speakers. However, for the male speakers, the F2 value is consistently higher for the Sydney speakers compared to the Hobartians. This suggests that the Hobart speakers may be realising the vowel in a slightly less fronted way. The F2 track for the Sydney female speakers contains some noticeable fluctuation. Several files have formants that impressionistically look somewhat inaccurate, and of these files, the formants generally seem to be lower than where they should be. These observations are reflected in the mean formant values taken at the initial point of measurement (15% interval) and the final point of measurement (85% interval), as indicated in Table 4.1.1.

Table 4.1.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the FACE vowel, separated by gender and location

	Sydney female	Hobart female	Sydney male	Hobart male
Initial F1	696 (σ 113)	765 (σ 112)	653 (σ 66)	655 (σ 93)
Initial F2	1994 (σ 492)	1955 (σ 234)	1778 (σ 232)	1616 (σ 56)
Final F1	427 (σ 34)	412 (σ 35)	339 (σ 27)	366 (σ 49)
Final F2	2481 (σ 437)	2551 (σ 183)	2338 (σ 196)	2165 (σ 86)
Duration	236 (σ 39)	272 (σ 43)	264 (σ 32)	231 (σ 29)

To determine the significance of any differences, using R, a linear mixed effects model was implemented using the package lme4 (Bates et al., 2015) for R (R Core Team, 2021). The independent variables are entered as an interaction between location and gender, with speaker as a random effect.

The independent variable for the first model was the initial F1 measurements, so the formula was F1_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.32$), so it was pruned to F1_15 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.34$), so it was pruned to F1_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p = 0.03$) for the initial F1 measurement of FACE, and not location.

The independent variable for the second model was the initial F2 measurements, so the formula was F2_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.46$), so it was pruned to F2_15 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.21$), so it was pruned to F2_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the initial F2 measurement of FACE, and not location.

The independent variable for the third model was the final F1 measurements, so the formula was F1_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.10$), so it was pruned to F1_85 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.56$), so it was pruned to F1_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the final F1 measurement of FACE, and not location.

The independent variable for the fourth and final model was the final F2 measurements, so the formula was F2_85 ~ location * gender + (1 |

speaker). The interaction between location and gender was not significant in this model ($p = 0.10$), so it was pruned to F2_85 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.39$), so it was pruned to F2_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the final F2 measurement of FACE, and not location.

4.2 PRICE vowel

The second vowel to be analysed is the PRICE vowel, which corresponds to /ɑe/ in the HCE system for transcribing Australian English (Harrington et al., 1997). These tokens were elicited using the prompt "hide". There are a total of 78 tokens in the sample for this vowel: 17 from Sydney females, 24 from Sydney males, 17 from Hobart females and 20 from Hobart males. Figure 4.2.1 illustrates the mean F1 and F2 trajectories of the vowel separated by location and gender, with Hobart speakers in blue and Sydney speakers in red.

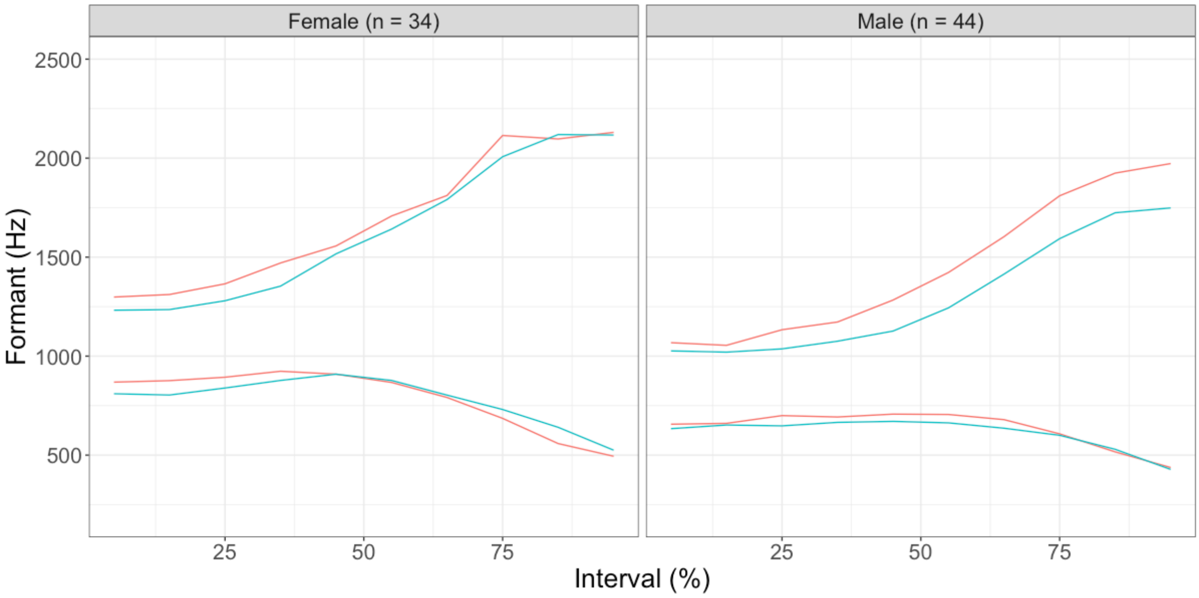

Figure 4.2.1 – Mean F1 and F2 trajectories for PRICE, separated by gender and location

There is no large visible difference in formant values between location for the female speakers. However, for the male speakers, the F2 value starts similar but then the Hobart speakers do not reach such a large final F2 value. This suggests that the Hobart speakers may be realising the vowel with a slightly less fronted second target. The F2 track for the Sydney female speakers once again contains some noticeable fluctuation. These observations are reflected in the mean formant values taken at the initial point of measurement (15% interval) and the final point of measurement (85% interval), as indicated in Table 4.2.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the PRICE vowel, separated by gender and locationTable 4.2.1.

Table 4.2.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the PRICE vowel, separated by gender and location

	Sydney female	Hobart female	Sydney male	Hobart male
Initial F1	876 (σ 92)	804 (σ 119)	661 (σ 76)	652 (σ 59)
Initial F2	1311 (σ 130)	1235 (σ 74)	1055 (σ 73)	1020 (σ 83)

Final F1	558 (σ 98)	641 (σ 58)	516 (σ 70)	530 (σ 54)
Final F2	2097 (σ 579)	2119 (σ 125)	1924 (σ 214)	1724 (σ 116)
Duration	282 (σ 39)	281 (σ 34)	295 (σ 42)	258 (σ 28)

A linear mixed-effects model was again implemented with an interaction between location and gender as the independent variables, and speaker as a random effect.

The independent variable for the first model was the initial F1 measurements, so the formula was F1_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.21$), so it was pruned to F1_15 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.14$), so it was pruned to F1_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the initial F1 measurement of PRICE, and not location.

The independent variable for the second model was the initial F2 measurements, so the formula was F2_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.53$), so it was pruned to location + gender + (1 | speaker). Location was not significant in this model ($p = 0.09$), so it was pruned to F2_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the initial F2 measurement of PRICE, and not location.

The independent variable for the third model was the final F1 measurements, so the formula was F1_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.09$), so it was pruned to F1_85 ~ location + gender + (1 | speaker). The final model shows that both location ($p \leq 0.05$) and gender ($p < 0.01$) are significant for the final F1 measurement of PRICE.

The independent variable for the fourth and final model was the final F2 measurements, so the formula was F2_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.24$), so it was pruned to F2_85 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.27$), so it was pruned to F2_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the final F2 measurement of PRICE, and not location.

4.3 GOAT vowel

The third vowel to be analysed is the GOAT vowel, which corresponds to /əʉ/ in the HCE system for transcribing Australian English (Harrington et al., 1997). These tokens were elicited using the prompt “hode”. There are a total of 77 tokens in the sample for this vowel: 18 from Sydney females, 21 from Sydney males, 18 from Hobart females and 20 from Hobart males. Figure 4.3.1 illustrates the mean F1 and F2 trajectories of the vowel separated by location and gender, with Hobart speakers in blue and Sydney speakers in red.

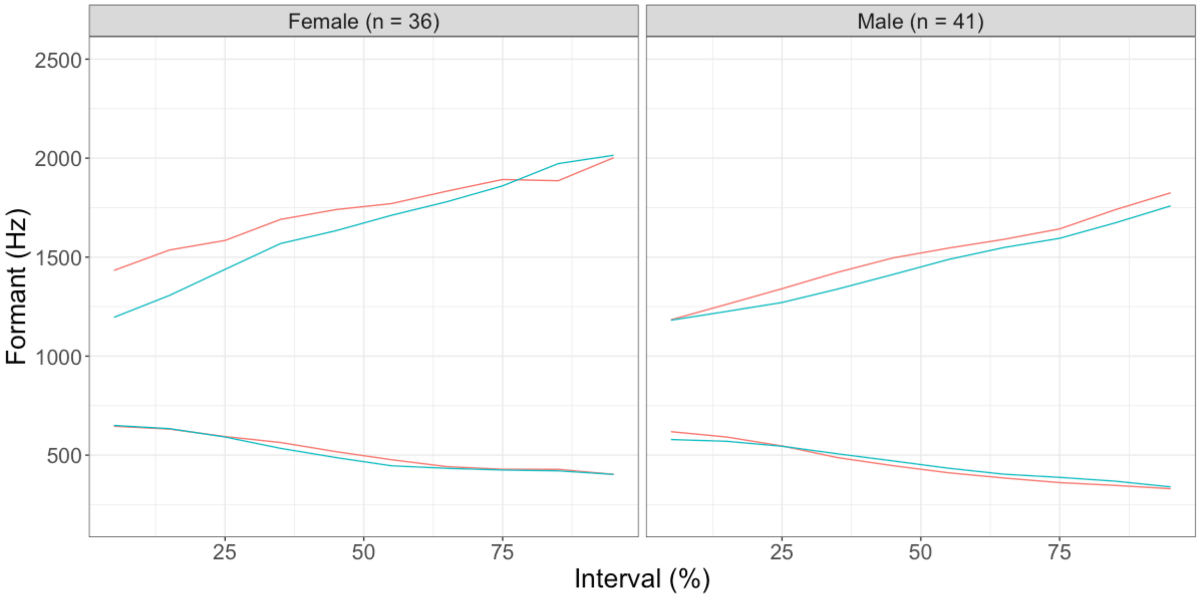

Figure 4.3.1 – Mean F1 and F2 trajectories for GOAT, separated by gender and location

There is no large visible difference in formant values between location for the male speakers. However, for the female speakers, the initial F1 value is higher for the Sydney speakers compared to the Hobartians. This suggests that Hobart speakers may be realising the vowel with a less fronted initial target. This can be observed in the mean formant values taken at the initial point of measurement (15% interval) and the final point of measurement (85% interval), as indicated in Table 4.3.1.

Table 4.3.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the GOAT vowel, separated by gender and location

	Sydney female	Hobart female	Sydney male	Hobart male
Initial F1	632 (σ 61)	634 (σ 67)	591 (σ 39)	570 (σ 28)
Initial F2	1536 (σ 146)	1307 (σ 64)	1261 (σ 96)	1226 (σ 90)
Final F1	429 (σ 39)	421 (σ 39)	347 (σ 34)	369 (σ 33)
Final F2	1886 (σ 278)	1972 (σ 181)	1739 (σ 169)	1673 (σ 102)

Duration	243 (σ 36)	280 (σ 45)	265 (σ 43)	244 (σ 22)

A linear mixed-effects model was again implemented with an interaction between location and gender as the independent variables, and speaker as a random effect.

The independent variable for the first model was the initial F1 measurements, so the formula was F1_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.56$), so it was pruned to F1_15 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.59$), so it was pruned to F1_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the initial F1 measurement of GOAT, and not location.

The independent variable for the second model was the initial F2 measurements, so the formula was F2_15 ~ location * gender + (1 | speaker). The interaction between location and gender is significant ($p < 0.01$). This is in line with the above observation that the Hobart have a noticeably lower initial F2 measurement for the GOAT vowel, but only for females.

The independent variable for the third model was the final F1 measurements, so the formula was F1_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.23$), so it was pruned to F1_85 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.49$), so it was pruned to F1_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the final F1 measurement of GOAT, and not location.

The independent variable for the fourth and final model was the final F2 measurements, so the formula was F2_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.26$), so it was pruned to F2_85 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.96$), so it was pruned to F2_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the final F2 measurement of GOAT, and not location.

4.4 MOUTH vowel

The fourth vowel to be analysed is the MOUTH vowel, which corresponds to /æɔ/ in the HCE system for transcribing Australian English (Harrington et al., 1997). These tokens were elicited using the prompt "howd". There are a total of 77 tokens in the sample for this vowel: 16 from Sydney females, 24 from Sydney males, 17 from Hobart females and 20 from Hobart males. Figure 4.4.1 illustrates the mean F1 and F2 trajectories of the vowel separated by location and gender, with Hobart speakers in blue and Sydney speakers in red.

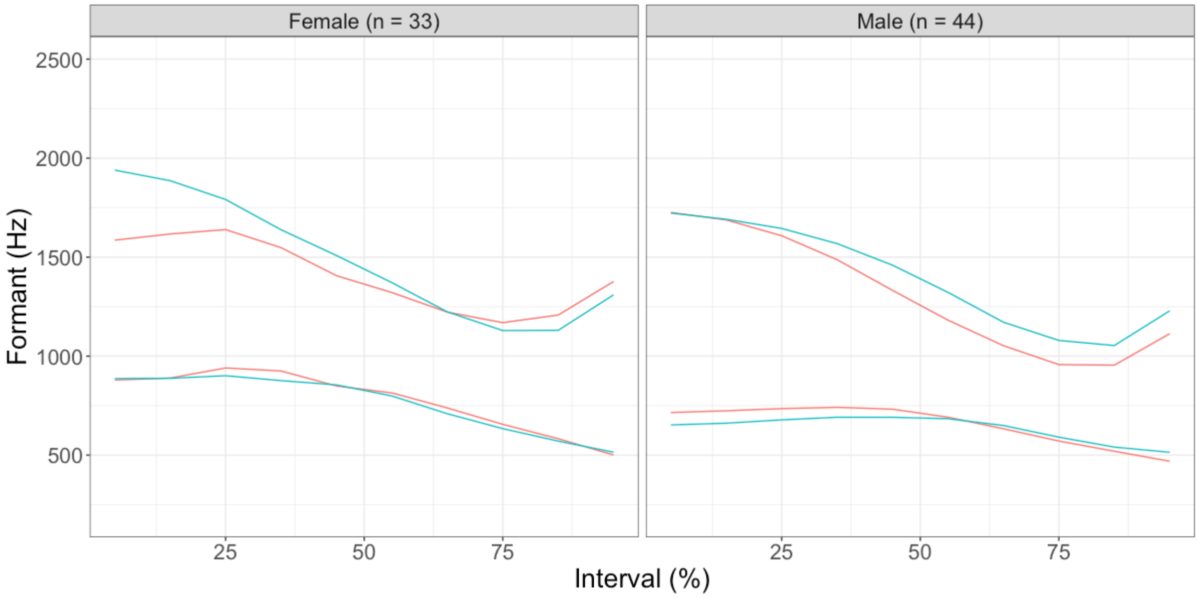

Figure 4.4.1 – Mean F1 and F2 trajectories for MOUTH, separated by gender and location

There is no large visible difference in the F1 values for either gender. However, for the females, the Sydney speakers appear to have a more compressed F2 trajectory compared to the Hobart speakers, with a lower initial F2 target frequency and a higher final F2 target frequency. For the males, the final F2 target appears to be higher for the Hobart speakers compared to the Sydney speakers, suggesting the final target is being pronounced slightly fronter. This can be observed in the mean formant values taken at the initial point of measurement (15% interval) and the final point of measurement (85% interval), as indicated in

Table 4.4.1.

Table 4.4.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the MOUTH vowel, separated by gender and location

	Sydney female	Hobart female	Sydney male	Hobart male
Initial F1	889 (σ 168)	888 (σ 92)	724 (σ 86)	661 (σ 89)
Initial F2	1617 (σ 331)	1886 (σ 143)	1688 (σ 158)	1691 (σ 122)
Final F1	583 (σ 129)	571 (σ 94)	520 (σ 75)	541 (σ 104)
Final F2	1208 (σ 155)	1130 (σ 142)	954 (σ 120)	1054 (σ 116)
Duration	274 (σ 33)	315 (σ 36)	301 (σ 48)	262 (σ 28)

A linear mixed-effects model was again implemented with an interaction between location and gender as the independent variables, and speaker as a random effect.

The independent variable for the first model was the initial F1 measurements, so the formula was F1_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.34$), so it was pruned to F1_15 ~ location + gender + (1 | speaker). Location was not significant in this model ($p = 0.47$), so it was pruned to F1_15 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor ($p < 0.01$) for the initial F1 measurement of MOUTH, and not location.

The independent variable for the second model was the initial F2 measurements, so the formula was F2_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.07$), so it was pruned to F2_15 ~ location + gender + (1 | speaker). Gender was the factor with the p value ($p = 0.37$) furthest above the significance threshold, so it was pruned to F2_15 ~ location + (1 | speaker). Location was still not significant in the final model ($p = 0.10$), so neither location nor gender are significant predictors of the initial F2 measurement for the MOUTH vowel in this sample.

The independent variable for the third model was the final F1 measurements, so the formula was F1_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.62$), so it was pruned to F1_85 ~ location + gender + (1 | speaker). Location was the factor with the p value ($p = 0.82$) furthest above the significance threshold, so it was pruned to F1_85 ~ gender + (1 | speaker). Gender was still not significant in the final model, so neither location nor gender are significant predictors of the final F1 measurement for the MOUTH vowel in this sample.

The independent variable for the fourth and final model was the final F2 measurements, so the formula was F2_85 ~ location * gender + (1 | speaker). The interaction between location and gender is significant ($p = 0.04$). This is in line with the previous observation that Hobart females have a slightly lower final F2 target value and Hobart males have a slightly higher final F2 measurement when compared to their Sydney counterparts.

4.5 CHOICE vowel

The fifth and final vowel to be analysed is the CHOICE vowel, which corresponds to /oɪ/ in the HCE system for transcribing Australian English (Harrington et al., 1997). These tokens were elicited using the prompt "hoyd". There are a total of 80 tokens in the sample for this vowel: 18 from Sydney females, 23 from Sydney males, 18 from Hobart females and 21 from Hobart males. Figure 4.5.1 illustrates the mean F1 and F2 trajectories of the vowel separated by location and gender, with Hobart speakers in blue and Sydney speakers in red.

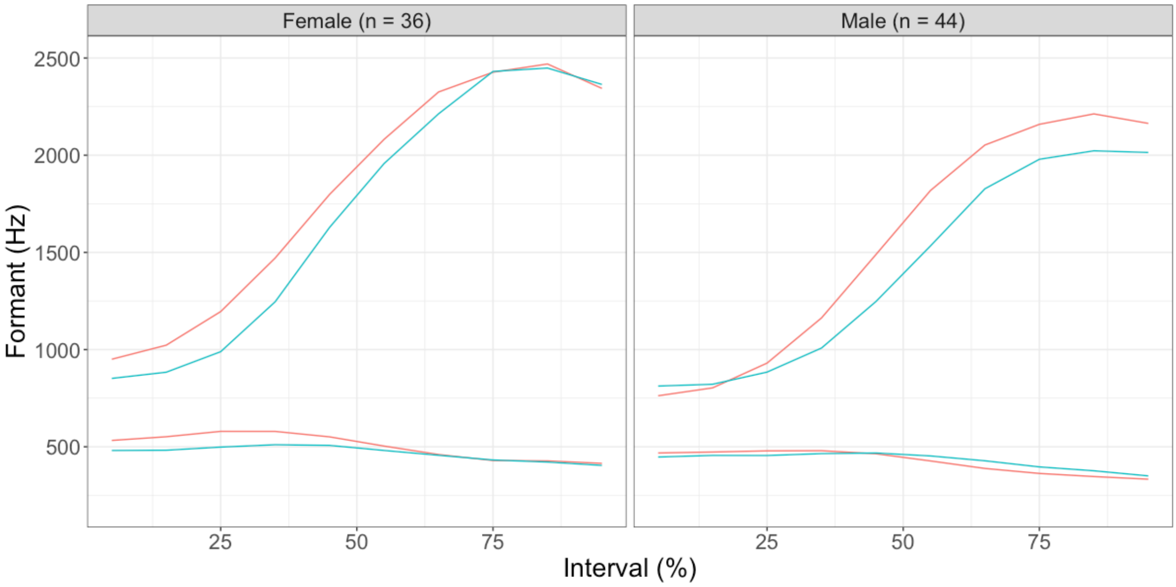

Figure 4.5.1 – Mean F1 and F2 trajectories for CHOICE, separated by gender and location

For the female speakers, the initial target formant values for both F1 and F2 appear to be slightly lower for the Hobart speakers. This suggests that the Hobart speakers may be realising the vowel in a slightly backer and closer way. The final targets, however, show little difference. For the male speakers, the initial target values also show little difference. However, the final F2 value is lower for the Hobart speakers. This suggests that the

Tasmanian speakers may be realising the final target in a slightly less fronted way. This can be observed in the mean formant values taken at the initial point of measurement (15% interval) and the final point of measurement (85% interval), as indicated in Table 4.5.1.

Table 4.5.1 – Mean formant values in Hertz and duration in milliseconds with standard deviations for the CHOICE vowel, separated by gender and location

	Sydney female	Hobart female	Sydney male	Hobart male
Initial F1	552 (σ 73)	482 (σ 30)	473 (σ 46)	456 (σ 28)
Initial F2	1023 (σ 101)	884 (σ 114)	803 (σ 95)	822 (σ 82)
Final F1	427 (σ 46)	421 (σ 29)	347 (σ 41)	376 (σ 37)
Final F2	2469 (σ 173)	2448 (σ 209)	2213 (σ 257)	2023 (σ 130)
Duration	244 (σ 31)	261 (σ 34)	248 (σ 41)	222 (σ 18)

A linear mixed-effects model was again implemented with an interaction between location and gender as the independent variables, and speaker as a random effect.

The independent variable for the first model was the initial F1 measurements, so the formula was F1_15 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model ($p = 0.14$), so it was pruned to F1_15 ~ location + gender + (1 | speaker). The final model shows that both location ($p = 0.2$) and gender ($p < 0.01$) are significant for the initial F1 measurement of CHOICE.

The independent variable for the second model was the initial F2 measurements, so the formula was F2_15 ~ location * gender + (1 |

speaker). The interaction between location and gender is significant (p = 0.04). This is in line with the previous observation that, for the CHOICE vowel, Hobart females have a slightly lower initial F2 measurement when compared to their Sydney counterparts, but there is little difference for the males.

The independent variable for the third model was the final F1 measurements, so the formula was F1_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model (p = 0.20), so it was pruned to F1_85 ~ location + gender + (1 | speaker). Location was not significant in this model (p = 0.31), so it was pruned to F1_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor (p < 0.01) for the final F1 measurement of CHOICE, and not location.

The independent variable for the fourth and final model was the final F2 measurements, so the formula was F2_85 ~ location * gender + (1 | speaker). The interaction between location and gender was not significant in this model (p = 0.25), so it was pruned to F2_85 ~ location + gender + (1 | speaker). Location was not significant in this model (p = 0.12), so it was pruned to F2_85 ~ gender + (1 | speaker). The final model shows that gender is the only significant factor (p < 0.01) for the initial F2 measurement of CHOICE, and not location.

5 Discussions and Conclusions

This study has examined to what extent extent there is variation in the way diphthongs are pronounced in Hobart, compared to in the more studied mainland location of Sydney. The analyses addressing this question show an interesting collection of results. The findings are highly specific to each vowel, with a unifying pattern hard to identify among the set of vowels. However, there are still interesting observations to be made that have been picked up by the models for several different diphthongs. In short, there is some limited variation between Hobart and Sydney in the pronunciation of diphthongs, but more research is needed to further unpack these differences and their nature.

The first vowel that was analysed was the FACE vowel, or /æɪ/ in the HCE system (Harrington et al., 1997). From the initial observations, a clear trend was visible in the data from the males, in which the F2 was consistently lower for the Hobart speakers, compared to the Sydney speakers. However, there was no significant interaction detected between location and gender in the linear mixed-effects models, even for the F2 measurements points at 15% and 85% into the vowel, and no effects detected for location by itself either after pruning. In addition to the possibility of the null hypothesis being true and the observed pattern is only due to random chance, it is possible that the sample size means that the models do not have sufficient power to detect any potential variation. In addition, the standard deviations, especially for the Sydney female speakers, are quite high, and, as discussed in chapter 4.1, there may be some formant tracking errors for this group of speakers. This means that the location trend that can be observed in the male speakers may also exist for the female speakers, but it could not be accessed with this data. It is also interesting to note that the possibility of this diphthong being generally backer for Hobart speakers aligns with Stanley's (2016) observations regarding the monophthongs of corresponding qualities by Hobart speakers. Both /æ/ and /ɪ/ were observed to be more back for Hobart speakers compared to Sydney speakers (Stanley, 2016), and the mean F2 value obtained from a sample of 3 male speakers from Hobart is 1644 Hz for /æ/ and 2175 Hz for /ɪ/ (Stanley, n.d.), both of which are very close Stanley's data and to what was observed at the 15% and 85% intervals respectively for the Hobart male speakers in this sample. For the first measurement, the Hobart male speakers in the present study had means of 655 Hz (σ 93 Hz) for F1 and 1616 Hz (σ 56 Hz) for F2, compared to the TRAP vowel /æ/ in Stanley's data, which showed means of 648 Hz for F1 and 1644 Hz for F2. For the second measurement, the present study showed a mean of 366 Hz (σ 49 Hz) for F1 and 2165 Hz (σ 86 Hz) for F2, compared to the KIT vowel /ɪ/ in Stanley's data, which showed means of 308 Hz for F1 and 2175 Hz for F2.

The second vowel that was analysed was the PRICE vowel, or /ɑe/ according to the HCE system (Harrington et al., 1997). From the initial observations, there was no obvious location-based difference for the female speakers, but for the male speakers, the F2 formants appear to diverge over the course of the vowel, being more similar at the 15% interval, with the Sydney speakers moving to higher F2 values by the 85% interval. However, this was again not detected using the linear mixed-effects models. The reasons for this may be any of the several previously described for the FACE vowel. It is possible, once again, that this observed trend is not a genuine pattern, only a product of random chance. It is also possible that the sample size too small to have adequate power to detect any location-based variation that may be present in the data. Once again, as discussed in chapter 4.2, there may also be some formant tracking errors here, particularly for the Sydney females. This could also have contributed to not detecting any effects, and a location effect of a lower final F2 value, and therefore a less fronted realisation of the vowel, may also be present for the female speakers in the same way it is visible in the data available for the males. This possible diphthong realisation difference for the PRICE vowel /ɑe/ also aligns with what Stanley found for the monophthong /e/, which was more back for the Hobart speakers compared to the Sydney speakers (Stanley, 2016).

The third vowel that was analysed was the GOAT vowel, or /əʉ/ according to the HCE system (Harrington et al., 1997). For this vowel, there were no differences between the locations that stood out for the male speakers, but, for the female speakers, the Hobart speakers had a noticeably lower F2 value at the 15% interval, suggesting a further back starting point for the diphthong trajectory. In this instance, the model did pick up that the interaction between location and gender was significant. This makes sense, given the difference was only noticeable on the graph for female speakers and not male speakers. There is no available data to compare with for how the schwa patterns as a monophthong for Hobart speakers. None of the other points tested showed any significant difference, both in terms of what was visible on the graph and what was calculated by the models. It is perhaps surprising that there is no difference for the final measurement, given the trend of GOOSE-fronting that has been observed patterning differently across various parts of Australia. This sound change concerns the vowel /ʉ/, which is approximately the second target of the GOAT diphthong /əʉ/. This sound change has progressed particularly strongly in New South Wales (e.g. Cox, 1998), to a greater extent than in other states (e.g. Butcher, 2006, 2012; Cox & Palethorpe, 2004). It was also observed that is has progressed further in New South Wales than in Tasmania (Stanley, 2016), as, in that study, Tasmania was linked more closely with Victoria's results. A link between the GOOSE-fronting phenomenon in monophthongs and a shift in the pronunciation of GOAT diphthong has been observed in New South Wales data (Cox & Palethorpe, 2001). However, when cross-referencing the diphthong data from the present study with Stanley's (2016) monophthong data, a relationship between how the second diphthong targets pattern between location and the corresponding monophthong data does not seem to be present in this data.

The fourth vowel that was analysed was the MOUTH vowel, or /æɔ/ according to the HCE system (Harrington et al., 1997). For this vowel, the male speakers appear to differ in F2 values at the 85% interval, and this time it is the Hobart speakers who have the higher F2 value, and therefore the more fronted realisation towards the end of the diphthong. For the female speakers, the Sydney sample appears to have a more compressed F2 trajectory than the Hobartians, with a lower initial F2 value and a higher final F2 value. In this particular case, duration is not a credible source for this difference, as, although the Sydney females have a shorter mean duration than the Hobart females, this difference is not remarkably large, and is of a similar magnitude to the difference that has been observed in some of the other vowels. The model for the initial F2 is particularly interesting as it is the only model in this study that does not pick up gender as having a significant effect on the formant values. This is likely to do with the compressed trajectory of the Sydney females resulting in an initial F2 value, that is in fact lower than the male values for either state. It is also interesting to note that both the interaction between location and gender and the standalone location variable came close to having a p-value below the threshold for statistical significance during the pruning process, although in neither case was it sufficiently low for the null hypothesis to be discarded. Perhaps a study with a larger sample size could more credibly provide an answer as to whether there exists a significant difference or not. The model for the F2 value at the 85%

measurement point, however, did pick up on a statistically significant interaction between location and gender in the sample. This reflects what can be observed in Figure Figure 4.4.1, as the difference for the male speakers and the female goes in opposite directions, relative to location. The difference for the male speakers here is particularly interesting, as it also goes against what has been previously shown in monophthong data. For the LOT monophthong /ɔ/, which is equivalent to the second target of the MOUTH diphthong /æɔ/, the Sydney speakers are the group with the fronter realisation, when compared to the Hobart sample (Stanley, 2016). It is unclear what causes the difference for the female speakers either, as previously mentioned, duration does not provide a credible explanation in this case.

The fifth and final vowel that was analysed was the CHOICE vowel, or /oɪ/ according to the HCE system (Harrington et al., 1997). For this vowel, there was an observable difference in the values of both formants for the female speakers at the 15% interval, with Hobart speakers having lower values for both formants, suggesting they pronounced the vowel at the initial measurement with a backer and less close realisation. Meanwhile, there was no noticeable difference for the male speakers in either formant at the 15% mark. However, by the 85% interval, any location-based differences were no longer apparent for the females, but, on the other hand, an F2 difference was observable for the males, with the Hobart speakers having a lower value, and therefore a less front realisation than the Sydney speakers. For this comparison, the model picked up significant influences of location, either alone or in an interaction with gender, for both of the formants at the 15% interval. F1 was found to have a significant effect for both location and gender separately, but not the interaction between them. This can be reconciled with the Hobart male speakers actually having a slightly higher F1 value than the Sydney males, which is a difference in the same direction as the more obvious difference observed in the females. For F2, it was the interaction between location and gender that had a significant effect. This makes sense, as the Hobart female speakers have a slightly higher F2 value than the Sydney females, but the Hobart males show no noticeable difference from their Sydney counterparts. However, no statistically significant location-based difference was found for the F2 difference at the 85% interval, even though that difference would have been in line with what has been shown by Stanley (2016) about the realisation of monophthongal KIT, or /ɪ/, being less fronted for Hobart speakers, compared to speakers form Sydney. Perhaps a future study with a larger sample size would be able to give a more credible analysis of whether the observed pattern is significant.

There are few unifying patterns to be observed that connect the observations in this data. One interesting point is that front targets such as [æ], [e] and especially [ɪ], often have lower F2 values for the Hobart speakers, suggesting a less fronted realisation, especially for the male speakers. This would be consistent with previous work on monophthongal realisations of the corresponding vowels, KIT or /ɪ/, DRESS or /e/ and TRAP or /æ/, by Hobart speakers (Stanley, 2016). Having observed this, it was not picked up as being influenced by location to a statistically significant degree in the case of the FACE, PRICE or CHOICE vowels. It is interesting that difference in F2 values not inclusive of the second target of the GOAT vowel, as GOOSE-fronting is a phenomenon known to be further advanced in New South Wales than other states, as previously discussed. However, a relationship between sound changes in Australian English monophthongs and diphthongs is not something that can be taken for granted. In their previously mentioned comparison of Sydney speakers over time, Cox and Palethorpe (2001) examined the diphthong changes with reference to the monophthong vowel space, but did not make any direct observations as to whether these were moving in tandem with each other. More research must be conducted into Tasmanian speakers' diphthongs, with larger sample sizes, to determine more credibly whether this observed effect is significant. It is interesting that previously observed trends regarding different orientation to region-based variation by gender (e.g. Billington, 2011; Cox & Palethorpe, 2019) has been replicated in this study. It is also clear that Tasmania has less distinctive realisations of diphthongs compared to what has been observed in other islands, relative to their mainland communities (e.g. Amos, 2011; Britain & Sudbury, 2010; Labov, 1963; Sutton, 1975). There could be several reasons for this, including the previously noted large culture of migration between

Tasmania and the mainland (Easthope & Gabriel, 2008). With a greater understanding of island identity and how it applies to Tasmania, a deeper understanding of why this is the case could be formed. It is possible that Tasmania is acting as a linguistically conservative region in some way, which is a trait that has been ascribed to the island previously (e.g. Wierzbicka, 2002).

There are some limitations in the analyses of some of the vowels. One recurring theme is data where differences in formant values may appear to be quite large, but the models do not pick up any significant location-based effect. This could be due to the fact that the study had a relatively small sample size. This does not lend the models enough power to be able to draw a credible conclusion about the nature of some of these differences and whether or not they are of interest. The limited amount of data also had an effect on the selection criteria for the study. With such a small sample to draw from, especially for the Hobart speakers, social factors such as residential history, education level and socioeconomic status could not be taken into account. This leaves any analysis vulnerable to influence from these factors. Future work in this area would benefit from a larger sample size, to increase the power of the models as well as provide room to include social factors other than location and gender into the analyses.

There are also some potential methodological limitations. 20 different linear mixed-effects models were built to analyse the data, with each formant being treated as a separate variable, and measurements taken at 15% and 85% analysed as if they had no relation to each other. This does not effectively deal with the reality that these are, in fact, all interconnected parts of the diphthong trajectory. A method that could be employed in future include taking the formant measurements with Euclidean distance, to measure the total difference across acoustic space, rather than isolating each formant as its own abstract variable. This will also reduce the total number of models employed in the study, which in turn reduces the total number of chances for Type I and II statistical errors, or chances that the null hypothesis is incorrectly either discarded or not discarded. Another potential source of methodological error is the formant tracks. Another type of model that could lend itself well to analysis of this type of data is generalised additive mixed models. The only processing of the formant tracks that was done was removing zero values. This is despite the fact that there remained a portion of the data that had formant values that differed greatly from what might be expected. This was especially true for the Sydney female speakers and could potentially be related to the use of creaky voicing. In any case, the issues with formant tracking also clearly impacted the efficacy of the linear mixed-effects modelling. The inaccurate formant tracks introduced a great deal of inconsistency into the data, which is at times reflected in the very high standard deviations for some of the formant statistics. Two particular cases of note are the F2 measurements at the 85% interval in the FACE and PRICE diphthongs. In both of these cases, it was clear that the Hobart speakers had a lower value than the Sydney speakers. However, both also exhibited a great deal of incoherence within the Sydney female speakers. The formant tracking problems brought down the mean of the Sydney female speakers, making it harder to differentiate between the location groups. The reduced consistency of the data also made the models less able to return a high degree of confidence in any effect, even if it is present. This no doubt contributed largely to no location-based effect being detected in either of these models. Although this problem only afflicted a relatively small subset of the data, future study would benefit from addressing this, whether by using a different algorithm altogether, or modifying the algorithm parameters for each speaker, personalised to the individual characteristics of their voice in a way that is more likely to yield accurate formant tracks for their specific case. Having a greater sample size will also reduce the impacts of a problem that affects just a specific group of speakers.

There are also many exciting avenues that can be explored in terms of further paths of enquiry. Alongside having more data for the cities already included in this analysis, it would be highly interesting to include more major mainland centres into the analysis in future. One city of particular interest would be Melbourne. As the closest major city to Tasmania, Melbourne is the most easily accessible city for Tasmanians to travel to. As previously mentioned, Tasmania also patterns with Victoria reasonably reliably in studies of lexical variation (e.g. Billington et al., 2015; Bryant, 1992; Kingstone, 2019) Therefore, it would be interesting to investigate the nature of any similarities or differences between Tasmanian speakers and Melbourne speakers in the pronunciation of diphthongs. Other mainland cities that have received some previous attention in sociophonetic research, such as Adelaide and Perth, may also serve as interesting points for further analysis.

The controlled wordlist data with the hVd frame also limits the generalisability of the results in this study. For the purposes of this study, it was not feasible to use natural speech data, as the time investment required for transcription to generate enough data for an analysis that takes into account all of the additional linguistic variables that using natural speech data introduces would have been very large. However, future study involving natural speech data will offer a more comprehensive picture of English as spoken in Tasmania.

Also, monophthongs have received considerable research interest over time in Australian English sociophonetics, and a wealth of knowledge has been built up around them that could be integrated into the conversation about diphthongs with some more investigation. Conducting more research into the extent to which monophthong changes and diphthong changes pattern with each other, both within and between speech communities, will deepen understanding by being able to integrate the body of work on monophthongs in Australian English for analysis of other speech features, such as diphthongs.

There is also room to explore further the relationship between different areas of Tasmania. As previously mentioned, there is traditionally a three-way regional divide on the island, split into the south (including Hobart), the north (including Launceston) and the north-west (including Burnie and Devonport). The Tasmanian data for this study only drew from southern Tasmania, and also only from an urban population. Although it is still considered regional for nearly all intents and purposes, Hobart is the most metropolitan city in Tasmania. Incorporating other regions of Tasmania, including rural areas, into future studies would create a richer picture of sociophonetic variation in English as spoken in Tasmania, and also help in properly understanding Tasmania's place in regional phonetic variation in Australian English.

Finally, just as it is not possible to assume that Tasmania is a homogenous entity in terms of pronunciation, it is also not possible to assume that all of regional Australia acts the same way either. There is a very wide range of places that are included in this concept of 'regional', each with different internal dynamics, as well as relations to different areas of the country. However, one thing that nearly all have in common is that they remain incredibly understudied. Much more work needs to be done using data from regional Australia in order to contribute to a greater understanding of regional phonetic variation in Australian English. One particular idea that would be interesting to investigate is whether regional Australia is genuinely linguistically conservative in relation to sound changes that have been observed as occurring in major metropolitan cities. Although it is a possibility, this idea negates the agency that regional areas have in their own language use, as well as their social networks. If the goal of regional speakers is to position themselves as similar to major metropolitan areas, there is no obvious reason for this to lag by considerable amounts of time, especially in the age of technology-enabled mass instant communication, where many, particularly younger people have social networks that expand beyond their local area and often involve people from large cities. An investigation is needed into to what extent regional Australia is linguistically conservative, and to what extent different parts of regional Australia are innovating on their own.

This study has revealed interesting findings about diphthongs of English as spoken in Tasmania. Through the linear mixed-effects modelling, it has shed light on the extent to which there is variation in the way diphthongs are pronounced in Hobart, compared to in the more studied mainland city of Sydney, including indications that there may be some interesting differences between the two locations. As the first sociophonetic study of diphthongs in Tasmania, it has expanded the breadth of literature on regional phonetic variation in Australian English. It has also given additional information from which an understanding of how Tasmania fits into the broader conversation about variation in Australian English. Investigation into the speech patterns of Hobart, as a regional centre, has also helped illustrate more about how linguistic variation operates in areas outside of Australia's major metropolitan cities, areas which have mostly received only limited attention in existing phonetic research. This study provides a good base for further investigation of the phonetics of English as spoken in Tasmania, and demonstrates the value of expanding the scope to beyond the major cities in the search for an understanding of region-based phonetic variation in Australia.

6 References

1270.0.55.005—Australian Statistical Geography Standard (ASGS): Volume 5—Remoteness Structure. (2016, July). Australian Bureau of Statistics. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.005

About AusTalk. (2012). AusTalk. https://austalk.edu.au/about/corpus/

Amos, J. (2011). *A Sociophonological Analysis of Mersea Island English: An investigation of the diphthongs /au/, /ai/ and /oi/* [PhD thesis, Rutgers University]. https://doi.org/doi:10.7282/T39C70NV

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *ArXiv E-Prints*, *arXiv:1406*. https://doi.org/10.18637/jss.v067.i01

Bernard, J. R. (1970). Toward the acoustic specification of Australian English. *STUF-Language Typology and Universals*, *23*(1–6), 113–128.

Billington, R. (2011). Location, Location, Location! Regional Characteristics and National Patterns of Change in the Vowels of Melbourne Adolescents. *Australian Journal of Linguistics*, *31*(3), 275–303. https://doi.org/10.1080/07268602.2011.598628

Billington, R., Gawne, L., Jepson, K., & Vaughan, J. (2015). *Mapping words around Australia*. The Linguistics Roadshow. https://bit.ly/AusWordsMaps

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.09 (15 February 2022)). http://www.praat.org/

Bradley, D. (1991). /Æ/ and /a:/in Australian English. In J. Cheshire (Ed.), *English around the world: Sociolinguistic perspectives* (pp. 227–234). Cambridge University Press. https://doi.org/10.1017/CBO9780511611889.016

Britain, D., & Sudbury, A. (2010). Falkland Islands English. In D. Schreier, P. Trudgill, E. W. Schneider, & J. P. Williams (Eds.), *The lesser-known varieties of English: An introduction* (p. 209). Cambridge University Press.

Bryant, P. (1992). *Regional variation in the lexicon of Australian English* [PhD thesis, Australian National University]. https://doi.org/10.25911/5d7a2638eeaaf

Butcher, A. (2006). Formant frequencies of/hVd/vowels in the speech of South Australian females. *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, 449–453.

Butcher, A. (2012). Changes in the formant frequencies of vowels in the speech of South Australian females 1945–2010. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 69–72.

Cassidy, S., Estival, D., Jones, T., Sefton, P., Burnham, D., & Burghold, J. (2014). The Alveo Virtual Laboratory: A Web Based Repository API. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.

Cox, F. (1998). The Bernard data revisited. *Australian Journal of Linguistics*, *18*(1), 29–55. https://doi.org/10.1080/07268609808599557

Cox, F. (2006). Australian English pronunciation into the 21st century. *Prospect*, *21*(1), 3–21.

Cox, F. (2006b). The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179. https://doi.org/10.1080/07268600600885494

Cox, F., & Fletcher, J. (2017). *Australian English pronunciation and transcription*. Cambridge University Press.

Cox, F., & Palethorpe, S. (2001). The changing face of Australian English vowels. In D. Blair & P. Collins (Eds.), *English in Australia* (Vol. 26, pp. 17–44). John Benjamins Publishing.

Cox, F., & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW/Victorian border. *Proceedings of the 2003 Conference of the Australian Linguistics Society*.

Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350. https://doi.org/10.1017/S0025100307003192

Cox, F., & Palethorpe, S. (2019). Vowel variation in a standard context across four major Australian cities. *Proceedings of the 19th International Congress of Phonetic Sciences*.

Crosswhite, K., Antoniou, M., & Doty, E. (2013). *TextGridMaker.Praat*. https://www.ling.upenn.edu/courses/Fall_2013/ling520/TextGridMaker.Praat

Dalla Fontana, M. (2019, February 6). *Is there such a thing as a Tasmanian accent?* ABC News. https://www.abc.net.au/news/2019-02-07/curious-hobart-is-there-a-tasmanian-accent/10780724

Docherty, G., Gonzalez, S., & Mitchell, N. (2015). Static vs dynamic perspectives on the realization of vowel nuclei in West Australian English. *Proceedings of the 18th International Conference of Phonetic Sciences*.

Docherty, G., Gonzalez, S., Mitchell, N., & Foulkes, P. (2019). An acoustic analysis of short front vowel realisations in the conversational style of young English speakers from Western Australia. *Proceedings of the 19th International Conference of Phonetic Sciences*, 1759–1763.

Easthope, H., & Gabriel, M. (2008). Turbulent Lives: Exploring the Cultural Meaning of Regional Youth Migration. *Geographical Research*, *46*(2), 172–182. https://doi.org/10.1111/j.1745-5871.2008.00508.x

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. https://doi.org/10.1121/1.4952387

Estival, D., Cassidy, S., Cox, F., & Burnham, D. (2014). AusTalk: An audio-visual corpus of Australian English. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*.

Ford, C., & Tabain, M. (2016). *Patterns of gender variation in the speech of primary school-aged children in Australian English: The case of /p t* *k/*. 16th Australasian International conference on Speech Science and Technology.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Grama, J., Travis, C. E., & Gonzalez, S. (2021). Chapter 13. Ethnic variation in real time: Change in Australian English diphthongs. In H. Van de Velde, N. H. Hilton, & R. Knooihuizen (Eds.), *Studies in Language Variation* (Vol. 25, pp. 292–314). John Benjamins Publishing Company. https://doi.org/10.1075/silv.25.13gra

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(2), 155–184. https://doi.org/10.1080/07268609708599550

Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, *36*(1), 89–103. https://doi.org/10.1146/annurev.anthro.34.081804.120633

Horvath, B. M. (1985). Variation in Australian English. The Sociolects of Sydney. *Cambridge Studies in Linguistics London*, *45*, 1–200.

Kingstone, S. (2019). *Mapping Australian English: An exploration of perceived and reported regional variation* [PhD thesis, Australian National University]. https://doi.org/10.25911/5d84aaeceabe2

Labov, W. (1963). The Social Motivation of a Sound Change. *WORD*, *19*(3), 273–309. https://doi.org/10.1080/00437956.1963.11659799

Ladefoged, P., & Johnson, K. (2014). *A Course in Phonetics*. Cengage.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.

Loakes, D., Hajek, J., Clothier, J., & Fletcher, J. (2014). Identifying /el/-/æl/: A comparison between two regional Australian towns. *15th Australasian International Conference on Speech Science and Technology*.

Loakes, D., Hajek, J., & Fletcher, J. (2017). Can you t[æ]ll I'm from M[æ]lbourne?: An overview of the dress and trap vowels before /l/ as a regional accent marker in Australian English. *English World-Wide. A Journal of Varieties of English*, *38*(1), 29–49. https://doi.org/10.1075/eww.38.1.03loa

Loakes, D., Hajek, J., Fletcher, J., & Clothier, J. (2018). *Variability in vowel categorisation behaviour amongst three diverse listener groups in Victoria, Australia*. 2018 Forum on Englishes in Australia. https://sites.google.com/view/auseng/archives/program-2018/abstracts/loakes-et-al

Metropolitan Boundaries. (2020). [Map]. New South Wales Department of Planning, Industry & Environment. https://www.planning.nsw.gov.au/-/media/Files/DPE/Maps/Metro-NSW-boundaries-map-2020-08-13.pdf?la=en

Mitchell, A. G. (1946). *The pronunciation of English in Australia*. Angus and Robertson.

Mitchell, A. G., & Delbridge, A. (1965). *The pronunciation of English in Australia*. Angus and Robertson.

Morrison, G. S., & Assman, P. F. (2013). *Vowel Inherent Spectral Change*. Springer. https://doi.org/10.1007/978-3-642-14209-3

Pawley, A. (2002). Using he and she for inanimate referents in English: Questions of grammar and world view. *Ethnosyntax: Explorations in Grammar and Culture*, 110–137.

Price, J. (2008). New news old news: A sociophonetic study of spoken Australian English in news broadcast speech. *AAA: Arbeiten Aus Anglistik Und Amerikanistik*, *33*(2), 285–310.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (Version 4.1.2 (1 November 2021)). Foundation for Statistical Computing. https://www.R-project.org

RStudio Team. (2021). *RStudio: Integrated Development Environment for R* (Version 2021.9.2.382 (2 September 2021)). RStudio PBC. http://www.rstudio.com/

Stanley, R. (n.d.). *An Analysis of Some of the Vowels of Australian English as Spoken in Melbourne, Sydney, and Hobart* [Unpublished Manuscript].

Stanley, R. (2016). Towards a Better Understanding of Regional Variation in Standard Australian English: Analysis and Comparison of Tasmanian English Monophthongs. *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*.

Stevens, M., Harrington, J., & Schiel, F. (2019). Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: A Journal of General Linguistics*, *4*(1). https://doi.org/10.5334/gjgl.620

Sutton, P. (1975). *Cape Barren English. Linguistic Communications: Working Papers of the Linguistic Society of Australia, No. 13.*

Travis, C. E., Grama, J., & Gonzalez, S. (In Progress). *Sydney Speaks Corpora*. Australian Research Council Centre of Excellence for the Dynamics of Language, Australian National University. http://www.dynamicsoflanguage.edu.au/sydney-speaks/

Vonwiller, J., Rogers, I., Cleirigh, C., & Lewis, W. (1995). Speaker and material selection for the Australian national database of spoken language. *Journal of Quantitative Linguistics*, *2*(3), 177–211. https://doi.org/10.1080/09296179508590049

Vonwiller, J., Rogers, I., Cleirigh, C., Taylor, B., Evans, G., Baldauf, R., Vaughan-Taylor, K., & Meares, P. (1998). *The Mitchell & Delbridge Tapes 1998 Re-issue, Companion Documents*. The University of Sydney. https://speech.library.sydney.edu.au/speech/MDReport.php

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468. https://doi.org/10.1121/1.427069

Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge University Press.

Wierzbicka, A. (2002). Sexism in grammar: The semantics of gender in Australian English. *Anthropological Linguistics*, *44*, 143–177.

Winkelmann, R. (2017). *PraatToFormants2AsspDataObj.R*. https://gist.github.com/raphywink/2512752a1efa56951f04

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, *45*, 392–410.