The relation between acoustic and articulatory variation in vowels: Data from American and Australian English

Arwen Blackwood Ximenes

Thesis submitted to the Western Sydney University for the degree of Masters of Philosophy MARCS

The MARCS Institute for Brain, Behaviour and Development

Western Sydney University

March, 2022

Principal Supervisor: **Professor Christopher Davis**

Associate Supervisors: **Associate Professor Jason Shaw**

Dr Chris Carignan

Dedication

I would like to dedicate this thesis with
Love to my Family.

To my dear husband, Fabiano, for your love and phenomenal support. And to my beautiful sons, Luca and Timon, who inspire me in so many ways.

Statement of authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

[redacted]

_____ 30/03/2022

Acknowledgements

I would like to sincerely thank my Masters Supervisors, all three of whom were in the role of my primary supervisor at some point. I count myself extremely lucky to have had the perfect primary supervisor for each stage of my thesis. Firstly, Associate Professor Jason Shaw who was an excellent mentor, providing the inspiration for this project and many hours of inspiring discussion about speech science. I am so grateful for all I learned from you in the time I worked and studied with you – from acoustic analysis to performing Electromagnetic Articulography and analysing articulatory data. My sincere thanks also go to Dr Chris Carignan, for the energy and enthusiasm you brought to my learning, and especially for the many hours patiently teaching me to code in R and Matlab to extract and analyse data, and to create figures, and also for all the engaging discussions. Last but not least to Professor Chris Davis, thank you for taking me on as a student when Jason and Chris left, and helping me with the thesis writing process. I am profoundly appreciative of your support and patience at this stage of my thesis, and life.

I would especially like to thank all the participants who agreed to be wired up so we could collect their speech data!

I am grateful to all my supervisors in various roles, as well as many colleagues past and present, and teachers, for inspiring and nurturing a love of speech science and for their support. Especially Professor Jonathan Harrington and Dr Robert Mannell, also Professor Felicity Cox, Dr Christine Kitamura and Dr Julie Vonwiller.

Fellow AHAA lab members. Especially Donald Derrick – for your help including technical expertise, mentoring, and encouragement. Mike Proctor, Cathi Best, Fellow AHAA lab students, Yassine Frej and Jia Ying.

MARCS Institute directors past and present, especially Professor Kate Stevens and MARCS HDR directors past and present: especially Professor Jeesun Kim (thank you Jeesun for your encouragement over the years). HDR admin support, Sonya, Craig, Jen, Krista, Jessica, Julia. Technical staff – Colin Schoknecht, Steven Fazio, Johnson Chen, Muawiyath Shujau, Donovan Govan, Dr Leidy Castro-Meneses. Also MARCS Institute admin staff past and present including – Darlene Williams, Gail Charlton, Karen McConachie.

A heartfelt thank you to my dear office buddies, Anne Dwyer, Sarah Wright, Peta Mills and Samra Alispahic. I miss you! And many other fellow students, too many to name, especially Stacey, Jaydene, Joey, Sarah F, Laurence, Tonya, Ruth, Valeria, and many more!

Many friends, especially Toni Cassisi, for programming support. Many other dear friends for moral support: Anna, Yas, Perry-jane, Niki, and others.

And lastly my family for their love and support. Mum and Dad for all your love, and your and support of my education and of our family over the years. My Aunty Lesley. My boys – Luca and Timon. You are my true inspiration. My husband, Fabiano – no words really can describe my gratitude for your love and enormous support and patience through everything.

List of Publications

Blackwood Ximenes, A., Shaw, J., & Carignan, C. (2016). Tongue positions corresponding to formant values in Australian English vowels. In *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology, 6-9 December 2016, Parramatta, Australia* (pp. 109-112).

Blackwood Ximenes, A., Shaw, J. A., & Carignan, C. (2017). A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America*, *142*(1), 363-377.

A note on Publications:

These papers were completed in collaboration with two of my supervisors at the time, Jason Shaw and Chris Carignan. Also note that the corpus of North American English and Australian English speech data had been collected prior to the commencement of this Masters degree. Blackwood Ximenes, Shaw, & Carignan (2017) constitutes the bulk of this thesis, but I completed the bulk of the work for this paper, being first author. For the data analysed in both papers listed above, I was also involved in much of the data collection. I was responsible for the method write up, data processing, data analysis, and producing figures and tables, with guidance and coding help from my supervisors. I wrote a good proportion of the introduction and discussion, however both supervisors also contributed significantly with writing, as we collaborated together. I have therefore used the pronoun "we" when referring to the published experimental work.

Table of Contents

List of tables

List of Figures and Illustrations

Abbreviations

AusE	Australian English
CVl	Consonant vowel /l/
dB	Decibel
EMA	Electromagnetic Articulography
F04	Female participant number four
*F*1	First formant
*F*2	Second formant
*F*3	Third formant
Hz	Hertz
IPA	International Phonetic Association
LL	Lower lip sensor
M01	Male participant number one
mm	millimeters
ms	milliseconds
MviewMVIEW	Multichannel visualisation application for displaying dynamic sensor movement
NAmE	North American English
NDI WAVE	Northern Digital Inc. WAVE
rms	Root mean square
SD	Standard deviation
TB	Tongue blade sensor
TD	Tongue dorsum sensor
TDx	Tongue dorsum sensor, horizontal dimension
TDz	Tongue dorsum sensor, vertical dimension
TT	Tongue tip sensor
UL	Upper lip sensor

Abstract

In studies of dialect variation, the articulatory nature of vowels is sometimes inferred from formant values using the following heuristic: $F1$ is inversely correlated with tongue height and $F2$ is inversely correlated with tongue backness. This study compared vowel formants and corresponding lingual articulation in two dialects of English, standard North American English and Australian English. Five speakers of North American English and four speakers of Australian English were recorded producing multiple repetitions of ten monophthongs embedded in the /sVd/ context. Simultaneous articulatory data were collected using electromagnetic articulography. Results show that there are significant correlations between tongue position and formants in the direction predicted by the heuristic but also that the relations implied by the heuristic break down under specific conditions. Articulatory vowel spaces, based on tongue dorsum (TD) position, and acoustic vowel spaces, based on formants, show systematic misalignment due in part to the influence of other articulatory factors, including lip rounding and tongue curvature on formant values. Incorporating these dimensions into our dialect comparison yields a richer description and a more robust understanding of how vowel formant patterns are reproduced within and across dialects.

Chapter 1. Introduction

1.1. Overview

Speech is arguably one of the most important modes of human communication. At its most basic, speech is sound produced by movements of the articulators (e.g., tongue, lips, jaw). Understanding the relationship between these articulatory movements and their resulting sounds is fundamental to speech science and has for many years been the focus of speech production research, remaining a strong area of interest and investigation today. This Masters thesis aims to contribute to this body of research through the comparison of articulatory and acoustic data both within but also across dialects. Dialects, although comprised of the same words, vary in terms of their phonetics, and therefore their acoustics and articulation, thus widening the scope of variation while keeping many factors constant (Foulkes, Scobbie & Watt (2010) Including this cross-dialectal aspect adds another dimension by which more light can be shone on the articulatory-acoustic relationship. One main aim of this thesis, therefore, is to characterise differences between two dialects, North American English (NAmE) and Australian English (AusE) with the overarching aim being to understand more about the articulatory-acoustic relationship.

One of the aims of dialect studies is to characterise differences between dialects. The majority of studies analysing phonetic variation across dialects have based their conclusions on differences in the acoustic properties of the dialects in question. Inferences about speech articulation made on the basis of acoustic analyses are often

useful in explaining patterns of variation across dialects and patterns of dialect change over time (Cheshire et al., 2011; Cox, 1999; Harrington et al., 2008). Although cross-dialect studies seldom compare dialects on the bases of both acoustic data and corresponding articulatory data directly, some studies do exist. For examples of ultrasound studies see the following: Scobbie et al., (2012) who report on both ultrasound and acoustic data of the GOOSE vowel in Scottish English; Turton (2017) who investigated cross-dialect /l/ vocalisation and darkening in the United Kingdom, and Kirkham and Wormald (2015) who studied articulatory variation of liquids between Anglo and Asian speakers of a dialect of British English. A small number of Electromagnetic articulography studies also exist, for example for a large study investigating tongue position in Dutch dialects see Wieling et al., 2016; also see Gorman and Kirkham (2020), who investigated the effects of coda consonants in dialects of British English. Nevertheless, dialect researchers still rely heavily on phonetic theory, based on the *Acoustic Theory of Speech Production* (Fant, 1960) — in particular, how acoustics relate to articulation — to bridge between readily available acoustic descriptions of dialect variation and speech articulation (what I shall be calling the acoustic method of speech research). One common assumption is that the first formant (*F*1) of a vowel is inversely correlated with tongue height; another is that the second formant (*F*2) of a vowel is inversely correlated with tongue backness.

This thesis assesses these oft assumed correspondences across two dialects of English: North American English (NAmE) and Australian English (AusE), reporting the tongue position of vowels and corresponding formant values for both dialects. One of its principal aims is to identify differences between dialects in both acoustic and articulatory

data, to determine whether the characterisation is complete looking at only acoustic information. In other words, does an approach relying wholly on established theory about how acoustics map to articulation, i.e. examining only acoustic data, fully capture variation in performance? Or, is it necessary to collect articulatory data to obtain a full characterisation of a language or of the difference between dialects? Another aim is to evaluate where the mapping seems to be robust and where it does not appear to capture the significant variation, and the reasons for this. To anticipate the conclusion, acoustic and articulatory descriptions reveal unique perspectives on how these dialects differ and offer examples of where typically assumed correspondences between formant values and tongue position break down.

The acoustic method of speech research remains the easiest method for investigation, being cheap, readily available, portable and non-intrusive. It was indeed also the only reliable method available to researchers for many years. However, due to advances in investigatory techniques, there is now suitable technology available to investigate the articulatory aspects of speech production. 3D Electromagnetic articulography (EMA) is one method which has been relatively recently developed and can achieve high spatio-temporal resolution of the movements of tongue, lips and jaw. Indeed, in Kochetov's (2020) general overview of current methods in articulatory phonetics research that surveys 379 published articles between January 2000 to December 2019 in prominent phonetic and phonological journals, EMA accounted for one third of all articulatory research in the twenty years preceding their paper. With this development of more precise articulatory investigation techniques, it has become possible to provide a direct comparison of the acoustics and articulation of a language, in order to look at the

relationship between acoustics and articulation. As a result, it is now of course also possible to compare the acoustics and articulation of dialects.

This project examines and compares parallel acoustic and EMA data from NAmE and AusE. Specifically, it seeks to answer whether the acoustics of AusE and NAmE monophthongs relate to the corresponding lingual and labial characteristics during speech in the manner predicted by the tube models espoused by Fant and others, which will be outlined in brief in section 1.3 below. Furthermore, it seeks to determine the conditions under which *F*1 and *F*2 are predictive of the articulatory properties of the monophthongs. Are there areas of the vowel space where examining acoustic data only is sufficient?

The remainder of the current chapter is designed as follows: Section 1.2 will present some relevant research on dialects in general and specifically on the two dialects in question. Section 1.3 aims to provide an overview of the relevant aspects of the *Acoustic Theory of Speech production*, touching on Source-filter Theory, tube models and Quantal Theory. Finally, section 1.4 will follow with an outline of the thesis aims, scope and research questions, ending with 1.5 an outline of the remaining three chapters.

1.2. Previous research

Here follows a brief account of the acoustics and articulation of the two dialects in question, NAmE and AusE.

The acoustics of NAmE vowels have been extensively reported (e.g., Hillenbrand et al., 1995; Peterson & Barney, 1952), and there are both studies focusing on vowel articulation

only (e.g., Johnson et al., 1993) and some that report both acoustic and articulatory data for a subset of vowels (e.g., Noiray et al., 2014).

Similar to NAmE, the acoustics of AusE vowels are well-studied (e.g., Harrington et al., 1997; Leung et al., 2020), but comparative articulatory data are lacking. Some recent studies on AusE vowel articulation focus on a small subset of AusE vowels. Tabain (2008) investigated the articulatory and acoustic properties of one vowel in different prosodic contexts. Watson et al., (1998) compared the acoustic and articulatory vowel spaces of AusE and New Zealand English (NZE). Their analysis covered four vowels, those in the words *hid*, *head*, *had*, and *herd*. Lin, Palethorpe and Cox (2012) looked at a larger number of AusE vowels in the /CVl/ context, although they focused on how vowel height influences lateral production (/CVl/) rather than on the phonetic properties of the vowels themselves.

The most comprehensive articulatory study of AusE vowels was undertaken over four decades ago (Bernard, 1970). Bernard reports on the results of an x-ray study investigating all the AusE vowels but does not report any quantitative measurements of the data. Bernard's qualitative description of x-ray data still constitutes the most comprehensive analysis of Australian vowel articulation to date in that it covers the entire vowel space, but due to technical limitations in synchronising acoustic and articulatory recording, the study does not report corresponding formant values. Thus, to date, dialect differences between Australian and American English are limited to those that can be inferred on the bases of acoustic measurements.

The known acoustic differences between NAmE and AusE dialects make for an intriguing test case of how reliably formant values reflect differences in articulation across dialects. To illustrate, consider the vowel referred to as the "GOOSE" vowel in Wells' (1982) lexical sets. The encroachment of GOOSE on front vowels, aka "GOOSE-fronting", has occurred in several dialects of English (Harrington et al., 2008; Watt & Tillotson, 2001; Scobbie et al., 2012; Cox, 1999). Increases in *F*2 may correspond to a more anterior tongue position, decreases in lip rounding (Harrington et al., 2011), changes in tongue curvature or pharyngeal cavity size, or some combination of these articulations. Comparison of dialects that differ in *F*2 values for GOOSE allows us to investigate the articulatory basis of a well-known acoustic difference between dialects. If the higher *F*2 observed in acoustic studies for the GOOSE vowel in AusE (see Cox, 1999 for AusE, c.f., Hillenbrand et al., 1995 for NAmE) is due to tongue configuration, we would expect the tongue to be more anterior for GOOSE in AusE speakers compared to NAmE speakers.

Another notable difference is the NURSE vowel. Reported formant values across dialects are substantially different for NURSE, which is rhotic in NAmE and non-rhotic in AusE. As with GOOSE, *F*2 for NURSE is higher in AusE than NAmE and, on the basis of *F*2 differences, is said to be more "front" in AusE (Cox, 1999).

Thus, both NURSE and GOOSE vowels have a higher *F*2 in AusE than in NAmE, but the articulatory basis of this formant difference, whether common or disparate for these two vowels, is not yet known. It is hoped that articulatory investigation of these phenomena may uncover the reason(s) for this difference.

1.3. Acoustic Theory of Speech Production

Given the overall aim of investigating the nature and extent of the acoustic-articulatory relationship, aspects of this thesis, as well as much of the research referred to above, are firmly rooted in Fant's *Acoustic Theory of Speech Production* (detailed in his pioneering work of 1960). This acoustic-articulatory research also draws on related theory such as Quantal Theory, as espoused by Stevens (1989). To achieve the aim of this thesis, the results of this study will be discussed primarily in relation to these speech production models. An understanding of the *Acoustic Theory of Speech Production* is therefore important in this endeavour, and it will be briefly described in this chapter.

These models are used by linguists and speech scientists to understand, or infer, what is happening on an articulatory level during speech. If the theory that links articulation and acoustics is shown to be essentially complete, then linguists could reliably make inferences about how production is made based on the acoustics. That is, we should be able to predict the articulatory properties of a sound based on changes in formant values. For example, a decrease in $F1$ would signal a raising of the tongue; similarly, a decrease in $F2$ would signal an increasingly posterior tongue position.

1.3.1. Evolution of the understanding of the relationship between acoustics and articulation of vowels

As outlined above, this research investigates the relationship between acoustics and articulation of vowels. Why vowels? They are particularly well-suited to investigating this relationship because they are relatively easy to measure both acoustically and articulatorily, being a sustained sound and vocal posture, at least for monophthongs.

There are also already well-developed models of how vocal tract shape influences the acoustics of vowels.

That there is a relationship between articulation and acoustics is now well established, however it was not always accepted. Vilain et al., (2015) describe how research on both aspects of speech production took some time to converge. The emphasis early on was on the articulation of vowels, rather than the acoustics. The study of acoustics followed when, in the late 19th Century, the understanding and technology to measure vowel acoustics were developed by Helmholtz. However, it wasn't until the mid-20th Century that the IPA articulatory quadrangle, depicting the relationship between articulation and the first two formants, *F*1 and *F*2, was described, and the tube models were developed. For early technological developments see: Hermann (1890), for early X-ray studies), Koenig, Dunn and Lacy (1946) and Potter, Kopp and Green (1947) for sound spectrograph development. Also, see Wood (1982) for a history on the development of the tongue arch model, where each vowel is defined in terms of tongue position (height and retraction), and on which the IPA vowel chart was based. These research methods have become more and more sophisticated, now allowing relatively precise parallel articulatory and acoustic data to be collected.

1.3.2. Source-filter theory

Before going into an explanation of specific tube models, however, other aspects of speech production need to be introduced, such as the source-filter theory of speech production, being an important basic concept. The *Acoustic Theory of Speech Production* is built upon this theory, so is therefore important to understand. According to the

Acoustic Theory of Speech Production, speech is comprised of the following physical properties giving rise to the acoustic signal:

1) a "**source**", created at the glottis, and which at its simplest is vocal fold vibration, caused by a combination of sustained subglottal pressure from the lungs and the physical properties of the vocal folds, and

2) a "**filter**", which modulates or filters the voice so that some of the frequencies produced at the vocal folds are attenuated but others are allowed to pass through. Thus at each vocal fold vibration a sound pressure wave is produced which is then filtered (or modulated) by the vocal tract above it, resulting in the vocal tract resonant frequencies. The resonant frequencies of the vocal tract are determined in part by the speed of sound in the body of air contained in the vocal tract, which is approximately 330 meters per second. However, more importantly for this thesis, they are also a result of supra-laryngeal vocal tract configuration, in other words vocal tract size and shape, which is modified by articulators during speech, resulting in a range of consonants and vowels which vary in their acoustic properties. The vocal tract resonances are difficult to measure, however they align closely with the formants, which are acoustic properties that can be analysed spectrographically. It is this "filter" component of the source-filter model that this research is looking at via the EMA data, predominantly three data points on the tongue, but also two at the lips, and comparing with formant data.

1.3.3. Fant's tube models

Fant's (1960) *Acoustic Theory of Speech Production* models vowels in terms of tubes. Fant proposed that the vocal tract above the larynx operated as a series of tubes varying on a number of parameters filtering the sound source in order to produce vowels. These were calculated initially using x-ray studies of Russian vowels. According to tube models, when a vocal tract changes size or shape it will resonate at different frequencies depending on the way in which it is modified by the articulators. Modifications of the vocal tract during speech production include varying the place of constriction, tongue height and vocal tract lengthening via lip rounding or larynx lowering.

1.3.3.1. Single tube model

The most basic tube model is the single tube with uniform cross-sectional area. The single tube can be used to model 'schwa' and central or 'neutral' vowels, such as in '*heard*', or in the case of this study, '*surd*'. In order to produce this vowel, the vocal tract is configured as a single tube with uniform cross-sectional area, closed at one end and open at the other (the glottis and lips respectively). This is the simplest vocal tract configuration for the purposes of calculating vocal tract resonances, and therefore the simplest way of understanding how these vocal tract resonances are generated.

The vocal tract resonances for a single tube model are directly related to the wavelengths of standing waves which are set up through wave propagation as a result of phonation during the vowel production. Each standing wave has a set wavelength for a tube of a given length and type. For the production of central vowels, the vocal tract most closely resembles a single tube open at one end at the lips and (effectively) closed at the

other end at the glottis. Knowing the speed of sound and the vocal tract length (and therefore the wavelength of the standing waves for each vocal tract resonance), it is possible to calculate the formant values predicted, as follows:

$$F1 = c / 4L$$

$$F2 = c / (4 / 3L)$$

$$F3 = c / (4 / 5L)$$

where c is the speed of sound and L is vocal tract length. As indicated above, the approximate speed of sound is 350 meters per second, and the average male vocal tract length is approximately 17.5 cm (0.175 m). Therefore, the predicted formants would be approximately 500 Hz, 1500 Hz and 2500 Hz, respectively.

Most vowel systems of the world's languages have more than three vowels. Calculating the resonant frequencies for other vowels requires consideration of the vocal tract as a series of tubes, as will be briefly described in the following section.

1.3.3.2. Two-, three- and four-tube models and Helmholtz resonators

Fant proposed that most vowels could be modelled using a two-tube model where the oral vocal tract was divided into a front and back tube. In this case, the maximum point of the tongue during constriction for a particular vowel constitutes the boundary between the two tubes. The two tubes, or "twin resonators" are considered closed at one end and open at the other (in other words they are quarter wave tubes), and can differ in cross-sectional area and length. Constriction location (tongue movement in the horizontal dimension), constriction degree (tongue height) and vocal tract length (modified via

larynx lowering or lip rounding) are the parameters which, when manipulated, affect the different resonating frequencies of the two tubes, and therefore also the formant measurements.

A brief description of the predictions of acoustic consequences on the first few formants of systematically manipulating the parameters of the two tubes follows. As a general rule, the cavity (tube) of greatest length gives rise to the lowest resonating frequency. This of course varies depending on the place of constriction (and therefore the relative lengths of the front and back tubes). If the back cavity is longer than the front cavity, it will be associated with the lowest resonating frequency, therefore it will give rise to *F*1 and the front cavity will give rise to *F*2. As the tongue moves in an anterior direction, *F*1 is predicted to rise and *F*2 is predicted to fall until the point where the front and back cavities are of equal length and *F*1 and *F*2 practically merge, in a phenomenon called 'acoustic coupling'. As the constriction location continues from back to front past this midpoint, it is the back cavity that now gives rise to *F*1 and the front cavity that gives rise to *F*2, with *F*1 and *F*2 predicted to continue in their aforementioned trajectories. *F*1 is also typically predicted to be negatively correlated with tongue height, which determines the cross-sectional area of each tube. Tongue height often co-varies with jaw opening (Lindblom & Sundberg, 1971), and a more open oral cavity (or front tube) with greater cross-sectional area is predicted to result in a lower resonating frequency. In addition, all vocal tract resonating frequencies are predicted to fall when the vocal tract is lengthened, such as by larynx lowering or lip rounding, for example in the case of GOOSE. The formant most commonly regarded as associated with lip rounding is *F*3.

Lindblom and Sundberg (1971) added to Fant's work with their articulatory model of speech production, based on their x-ray studies, contributing to our understanding of the role of the jaw in achieving articulatory goals. Their fixed jaw studies showed that compensatory tongue shapes were formed to keep the expected acoustics of vowels constant, and they proposed the idea of "articulatory synergism", where jaw position is optimised to reduce the need for tongue deformation. Such x-ray studies, whilst having contributed much to our understanding of speech production, are lacking in detailed measurements. As alluded to earlier, much more precise measurements can now be taken for example using point-tracking methods. A number of studies have employed these methods using for example EMA or X-Ray Microbeam data to investigate acoustic-articulatory relations in vowels with varied success. However, the picture is still incomplete. For example, one study extracted data from only one magnet on the tongue tip in order to investigate diphthongs (Dromey, Jang, & Hollis, 2013). Another looked at variability in acoustics versus articulation and argued that their findings were evidence for a strong relationship between acoustics and articulation. This may be the case, however this doesn't quantify the relationship *per se* (Whalen, Chen, Tiede, & Nam, 2018). Gorman and Kirkham (2020) more recently used EMA to investigate the acoustic-articulatory relations through comparing two dialects of English, and made some interesting findings, noting some exceptions, or unexpected mismatches. This work will be referred to in more detail in Chapter 4, however it is important to note here that data from only two different vowels was analysed. Thus more definition is needed, and despite these and other studies, some of which I have already referred to earlier, there remains a shortage of data.

The calculations involved in determining the formants with a two-tube model are more complex than those for a single tube model. However, Stevens (1989) suggested that some vowels, such as non-low front vowels may be more usefully modelled using an even more complicated three-tube model, where the front and back cavities are separated by a smaller, narrower tube (at the point of maximum constriction) (Stevens, 1989). This creates a Helmholtz resonator with the back cavity, which has the effect of lowering the resonating frequency of the back cavity, and is associated with *F*1 (Stevens, 1989). Urbassek (2014) was able to demonstrate a fairly reliable two-tube model for 'schwa' and the 'corner' vowels [iː], [aː], and [uː], but argued that using a three-tube model would provide greater accuracy across the whole vowel space, especially for vowels in the sequence [uː] – [oː] – [ɔː] – [aː]. A four-tube model would provide even more accuracy, however the increase in complexity of the calculations makes this far less practical.

1.3.4. Quantal Theory

One theory that was proposed subsequent to the earlier x-ray study work and that provides an important *explanans* for this thesis is Quantal Theory (Stevens, 1989). This theory provides a more nuanced view of the tube models presented by Fant. It proposes that there are alternating regions of stability and instability in articulation of vowels. In the stable regions, relatively small movements of the articulators result in relatively large changes in acoustics. Conversely, in regions of stability, where a relatively large variation in articulator placement does not result in much acoustic change would allow speakers some "lee-way" in terms of accuracy of vowel articulation, which could be argued to be important in connected speech in particular. It is also proposed that the stable regions occur in areas of convergence as described above, i.e. where 'acoustic coupling' exists.

Stevens describes this relationship as being non-linear in nature, i.e. acoustic variations will be smaller or larger depending whether the articulator movements occur within a region of stability or not.

1.4. The thesis

1.4.1. Study aim and scope

This thesis sets out to address the lack of parallel articulatory and acoustic data (in general, but for AusE in particular). It also attempts to demonstrate the extent of the assumed relationship between articulation, especially lingual articulation, and acoustics, by comparing the acoustic and articulatory properties of vowels. This thesis assesses these assumed acoustic and articulatory correspondences (based on phonetic theory) across two dialects of English.

I report on the tongue position and lip rounding data of ten monophthong targets and corresponding first two formant values for two dialects of English, North American English (NAmE) and Australian English (AusE). Vowel targets are defined as the point of minimum velocity of the tongue dorsum sensor (TD), thus the data is static in nature. I then present a discussion of how the dialects differ both in terms of acoustics and articulation, and where in the data assumed correspondences between formant values and tongue positions break down.

1.4.2. Research questions

The overarching research question of this thesis is: *What is the relationship between the acoustics and articulation of vowels?* As noted above, the scope of this thesis is,

however, limited to monophthongs from two dialects of English and measuring mainly the lingual position at a single time point (the vowel target defined by the minimum velocity of the TD using EMA) and the first two formants at that same time point.

In order to address the main research question, the following sub-questions were also posed (for which I provide tentative answers or describe what the nature of the answer will be):

1) *What are the lingual and labial characteristics of Australian English and North American English monophthongs?*

   The results for this question will be descriptive in nature.

2) *How do these articulations relate to the acoustics?*

   a) *Under what conditions is F1 predictive of tongue position?*

      The hypothesis is that the prediction is likely to hold.

   b) *Under what conditions is F2 predictive of tongue position?*

      The hypothesis is that the prediction is likely to hold in general but that the relationship may break down under certain conditions ***

3) *What can articulatory data contribute to our understanding about vowels that is not revealed by acoustic data?*

   The answer to this question will be descriptive in nature.

4) *Will there be variation in the degree to which individual speakers' articulatory and acoustic data are correlated?* Prediction: yes

1.4.3. Thesis organisation

This thesis is organised as follows. At the beginning of this chapter, I outlined the aims of the thesis. I then presented some relevant research on dialects in general and specifically on the two dialects in question, and gave an overview of vowel production (including tube model theories) and relevant theories of speech production. In Chapter 2 I describe the methods used for the acoustic and articulatory investigation in detail. In Chapter 3 I present acoustic (*F*1 and *F*2) and articulatory results (TD position and measures of lip rounding) separately for each dialect. Results of acoustic-articulatory correlations are then presented, as well as a comparison of the two dialects using linear mixed-effects models. Chapter 4 is a discussion of the results in light of the common assumptions based on phonetic theory, the predictions made by the tube models, and existing research. I first follow up the correspondences between acoustics and articulation, including individual data, and then include a discussion of how conclusions about how vowels differ across dialects vary depending on the type of data examined (articulatory or acoustic). I finish with limitations of the present study, suggestions for future research with an account of a preliminary investigation, and finally, concluding remarks.

Chapter 2. Methods

Articulatory and acoustic data were collected as part of a larger EMA study at the MARCS Institute, Western Sydney University. Five speakers of NAmE and four speakers of AusE were recorded producing multiple repetitions of ten monophthongs embedded in the /sVd/ context. Articulatory data relating to vowel targets, specifically velocity minima, were extracted with corresponding acoustic data. A description of the methods follows, with further explanation of various aspects where necessary. Although a standard approach to EMA was used, the use of the velocity minimum to define the vowel target is not as common, so a rationale for this is also included.

2.1. Subjects

All participants were recruited via the experimenters' professional and social networks by word of mouth (all except one of the participants were sourced from the Western Sydney University community). Word of mouth (or "snowballing") as a subject recruitment technique has practical and methodological advantages which outweigh issues such as sampling bias (Holmes & Hazen, 2013). Natural consequences of this recruitment method include increased likelihood of motivation and trust. This leads to a greater likelihood of eliciting "naturalistic" speech. Given the invasive nature of EMA, it was important that subjects felt safe and could trust the researchers, so having a connection in some way was helpful. In addition, only subjects who genuinely wanted to contribute to speech research were likely to volunteer, again given the invasiveness of

EMA. It was also necessary to use researchers' networks to target NAmE speakers, who are a small subset of the general population in Australia.

Data were analysed from five NAmE speakers (three females) and four AusE speakers (two females). The former group of speakers range in age at time of recording from 31 to 60 and the latter range in age from 20 to 42. All of the speakers were residents of the Greater Sydney region. Three of the North American speakers had lived in Australia for less than two years. The other two speakers had lived in Australia for 8 years and 14 years, respectively, at the time of recording. The NAmE speakers originated from diverse regions of North America as follows: F04 (California), F10 (Chicago), F11 (New England), M01 (Nova Scotia), and M02 (Washington State). The AusE speakers all originated from the state of New South Wales, and while there is some linguistic diversity in New South Wales (including the existence of sociolects in Western Sydney, whence two of the speakers originated), there is significantly less variation in this group than in the NAmE group. Reasons for this include geographical distribution and sociolinguistic profiles.

2.2. Materials

Stimuli comprised a list of lexical items and nonce words containing 15 vowels, including 10 monophthongs, in the sVd context. This paper focuses on analysis of the monophthongs. The stimulus items are provided in Table 1. Alongside the orthographic stimuli (column 1), we provide the IPA symbol corresponding to the vowel in North American and Australian English and the reference word, or "lexical set", for the vowel devised by Wells (1982). The reference words disambiguate the spelling, which is particularly useful for nonce words and were used as a guide for participants to produce

nonce stimuli with the intended target vowel. This set of monophthongs covers the whole of the NAmE and AusE acoustic vowel spaces. The only monophthong missing is START from AusE, which according to Cox (2006), does not differ in its formant structure from the AusE STRUT vowel, the difference being only in length. As indicated by the NAmE IPA symbols in Table 1, a merger between THOUGHT and LOT was expected for some NAmE speakers given the diverse regions of origin.

Table 1. List of materials for North American English (NAmE) and Australian English (AusE).

sVd stimuli	Australian English vowels (IPA symbols)	North American English vowels (IPA symbols)	Lexical Set
sad	æ	æ	TRAP
said	e	ɛ	DRESS
sawed	oː	ɔ (ɑ)	THOUGHT
seed	iː	i	FLEECE
sid	ɪ	ɪ	KIT
sod	ɔ	ɑ	LOT
sood	ʊ	ʊ	FOOT
sud	ɐ	ʌ	STRUT
sued	ʉː	u	GOOSE
surd	ɜː	ɝ	NURSE

2.3. Procedure

The movements of speech articulators were tracked using a Northern Digital Inc. Wave EMA system at a sampling rate of 100 Hz. This system uses an electromagnetic field to track the movement of small receiver coils or sensors (~3mm in size) glued or taped to the articulators. The electromagnetic field induces an alternating current in the sensors, and the strength of this current is used to determine the position of the sensors in relation to the transmitter. Articulatory movements are captured in the vertical, horizontal and

lateral dimensions with high spatial resolution (< 0.5mm rms error; Berry, 2011). In this study, we focused on movements in the horizontal and vertical dimension, since these are the dimensions typically assumed to correspond to values of the first two formants. The sensor trajectories were synchronised to the audio signal during recording by the NDI system. EMA sensors were glued to the following articulators along the midsagittal plane: jaw, below the lower left incisor; lips, at the vermillion edge of the upper (UL) and lower lip (LL); tongue tip (TT), tongue blade (TB) and tongue dorsum (TD). The TD sensor was placed as far back as comfortable for the participant. The TT sensor was placed approximately 5 mm back from the tongue tip and the TB sensor was placed midway between the TT and TD sensors. The three lingual sensors and the UL sensor (with tape) can be seen in Figure 1, with connecting wires. The wires attached to the LL and jaw sensors can also be seen below the tongue. Speech acoustics were recorded using a shotgun microphone at a sampling rate of 22050 Hz.

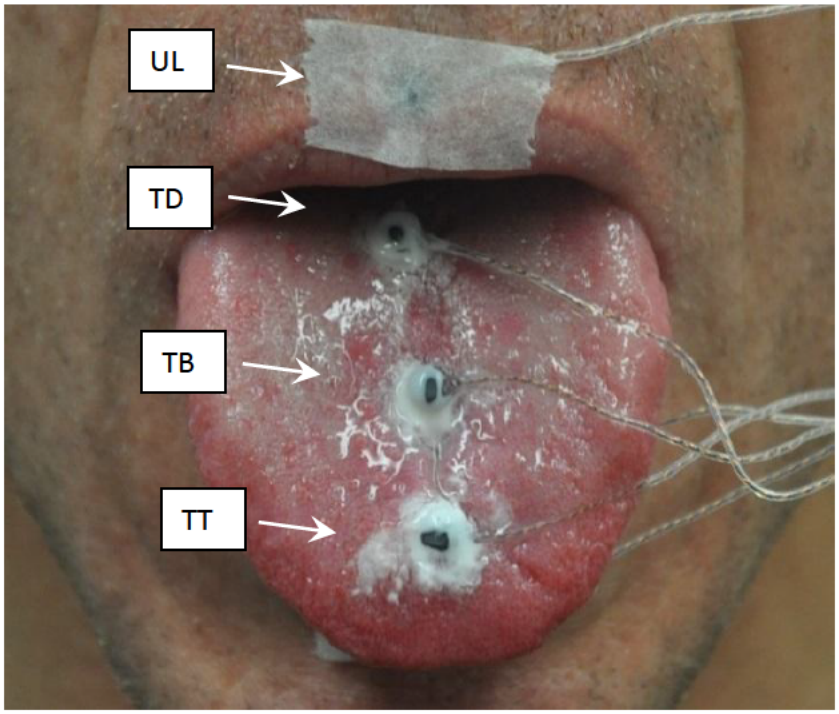


Figure 1. Image of protruded tongue with labelled sensors. UL = upper lip, TT = tongue tip, TB = tongue body, and TD = tongue dorsum. The wires for lower lip and jaw sensors are also visible.

The target stimulus words were displayed on a computer monitor placed outside of the magnetic field (with a volume of 300 mm^3). One word was presented per trial. There were 15 trials (one per vowel) per block and eight blocks in the experiment. This resulted in 15 (vowels) x 8 (repetitions) = 120 vowel tokens per participant. Of the recorded data, the monophthongs consist of 10 (vowels) x 8 (repetitions) = 80 tokens per participant, 320 monophthong tokens in total for AusE (four speakers). The stimulus presentation order was uniform across blocks. As there were five NAmE speakers, 400 monophthongs would have been recorded; however, half of one male speaker's session was not recorded successfully (due to problems with sensor adhesion). Accordingly, only 360 tokens were recorded in total for NAmE. There was an error in the audio for the first twenty tokens of one female AusE speaker, so only 300 tokens were recorded for AusE. Other technical problems due to data acquisition, analysis, and mispronunciation, resulted in seven more

tokens out of 660 total across accents (~4% of the data) being excluded from the analysis: four tokens of NAmE and three tokens of AusE.

Head movements were corrected using custom written MATLAB functions developed by Mark Tiede and revised by Donald Derrick. Sensors taped to the nasion and left/right mastoid processes were used as stable reference points for the head correction procedure. The articulatory data were rotated relative to the occlusal plane so that the origin of the coordinate system corresponds to a point immediately posterior to the incisors. The occlusal plane was established by having the participant bite down on a protractor with 3 sensors affixed in a triangular formation.

The NDI Wave system has an automatic head-correction procedure. This was accidentally applied during recording of two of our NAmE speakers, rendering the data relative to the right mastoid sensor. After rotating this data to the occlusal plane, the location of the sensors within our reconstructed coordinate system differed systematically from the other speakers. The normalisation step described below in Section 2.6 rendered all data relative to the centre of the articulatory space, correcting for differences introduced in post-processing.

2.4. Articulatory measurements

Figure 2 shows a token of the word *seed*, which is used here to illustrate the measurement procedure. The topmost pane shows the tangential velocity of the TD sensor, based on movements in vertical and horizontal dimensions, the middle pane shows the TD trajectory in the vertical dimension and the lower pane shows the speech

waveform. The three panes are synchronised in time. Vertical dashed lines indicate the velocity peaks associated with movement towards and movement away from the vowel target. The solid vertical line indicates the velocity minimum which occurs for this vowel at the highest point reached by the TD.

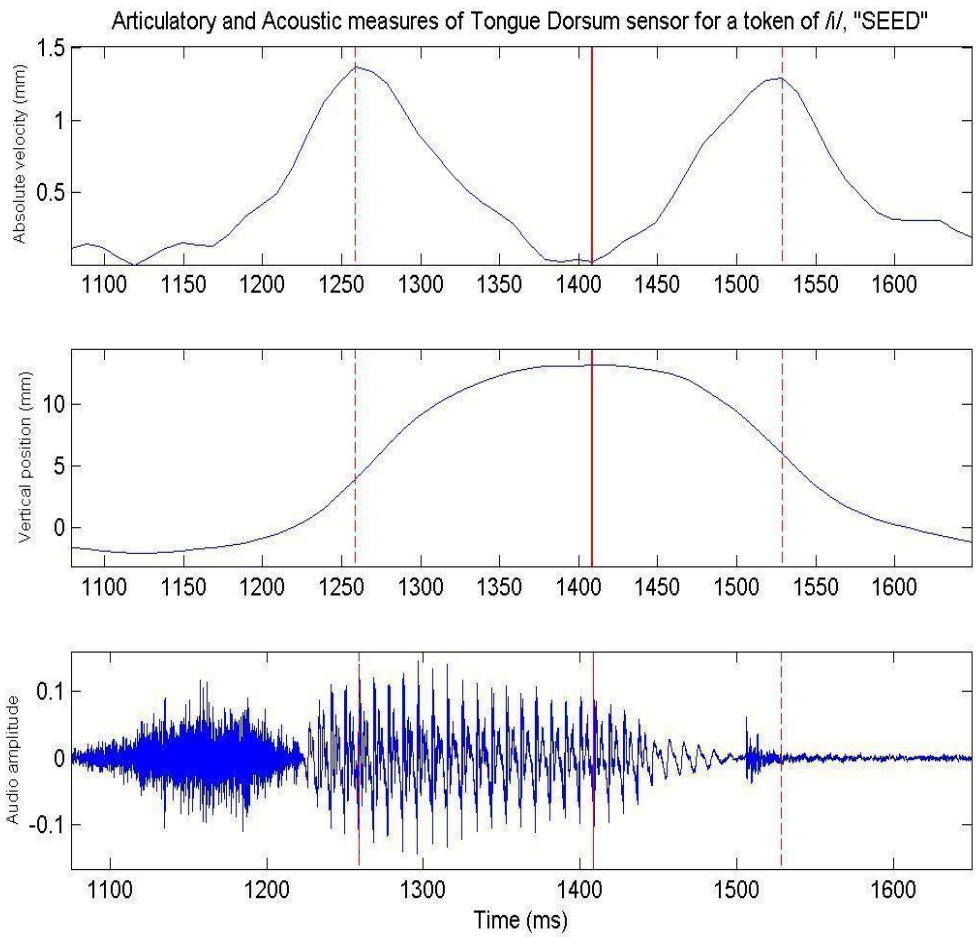


Figure 2. Labelling procedure for a "seed" (FLEECE) token. The lower pane contains the speech waveform. The lower pane contains the speech waveform. The middle pane represents the trajectory of the tongue dorsum sensor in the vertical dimension (the occlusal plane was set to 0 mm). The upper pane represents the velocity of the tongue dorsum sensor. Vowel target in all three panes is indicated by a solid vertical line. Velocity peaks in movements toward and away from target are indicated by dashed lines.

We determined the vowel target based on this velocity minimum. Measurements were extracted from sensor trajectories in the vertical and horizontal dimensions based on

timestamps labelled using *findgest*, an algorithm developed for the MATLAB-based software package, "Multi-channel visualisation application for displaying dynamic sensor movement" (Mview), by Mark Tiede at Haskin Laboratories. This program was used to detect the nearest tangential velocity minimum of the TD sensor during the interval corresponding to the vowel. We then extracted positional coordinates from all the lingual sensors and from the LL and UL sensors at this vowel target landmark.

As mentioned earlier, this is not the most common method of defining the vowel target as often the articulatory data is extracted at the time point of an acoustically defined target. However, other studies (e.g. Gafos et al., 2010; Ratko et al., 2016; Tilsen and Goldstein, 2012) have used velocity profiles to determine articulatory or "gestural" landmarks. This is consistent with the understanding of the articulation of speech sounds as made up of discrete functional articulatory gestures, as described by Task Dynamics (Saltzman & Munhall, 1989). Within this framework, segments or phonological units are viewed as continuous movements or gestures, which also may have some overlap between consecutive units. This study takes one point in time, the time point of maximum constriction of the articulator (here, the point of minimum velocity of the TD sensor), in order to provide a snapshot of the articulatory space which can then be viewed in relation to the acoustic space.

For some tokens, the point of minimum velocity in the TD trajectory did not give a reliable indication of the vowel target. This was the case, in particular, for vowels with a long period of little or no movement, i.e., a quasi-steady state. In these tokens, since velocity remains relatively constant, selecting the vowel target based on an absolute

velocity minimum is somewhat arbitrary. When the time point of minimum velocity in the TB sensor trajectory provided a clearer indication of the vowel target than the TD sensor, we extracted articulatory coordinates from the minimum velocity of the TB sensor instead of the TD sensor.

2.5. Acoustic measurements

Formant listings (*F*1 and *F*2) were extracted using LPC analysis in PRAAT (Burg method with a 25 millisecond window length and a 6 dB per octave pre-emphasis from 50 Hz) at the point determined by the minimum velocity of the TD (see, e.g., Shaw et al., 2013: 166-167). Results were then inspected visually, and outliers were hand corrected as needed. Using the time points extracted from the articulatory measures for the acoustic analysis enables a direct comparison between articulation and acoustics. Parsing vowel targets using the point of minimum velocity in the articulatory data follows similar general principles used to identify formants in Cox (2006) and Harrington *et al.* (1997), where vowel targets were identified based on formant displacement patterns, e.g., max/min *F*1/*F*2, depending on vowel. Max/min formant values relate closely to the minimum velocity of articulator movement in our data. Other acoustic studies have used the acoustic midpoint of the vowel, which did not correspond as consistently to the velocity minimum of the TD or TB sensors in this data, as can be seen, for example, in Figure 2, where the velocity minimum occurs well after the midpoint of periodic energy in the acoustic signal.

2.6. Analysis

One of the challenges of analysing speech production across speakers is that anatomical differences influence both the formant values and EMA positional coordinates. In the case of formants, differences in vocal tract length influence the average formant values. In articulatory data, differences in tongue shape, volume, and sensor placement lead to different average values across speakers. For example, a retraction of the TD to a point 30 mm behind the front teeth would have a different meaning between speakers due to variation in tongue size. In both cases, because of differences in anatomy, between-speaker differences for the same vowel can be larger than within-speaker differences across vowels. In order to facilitate comparison across our speakers, we normalised both the formant values and the lip and tongue positional coordinates by calculating *z*-scores of sensor positions and formant values, a method established by Lobanov (1971) for vowel formants and extended to EMA sensor positions (e.g. Shaw et al., 2016). Sensor positions were normalised 1) across the three lingual sensors, TD, TB and TT and 2) across the two labial sensors, UL and LL. The horizontal and vertical dimensions were normalised separately. To provide a measure of lip protrusion, we calculated the mean horizontal position of the UL and LL sensors. Normalisation preserves the within-speaker structure of the data, but allows for a direct comparison across speakers, serving the goal of dialect comparison.

Due to the issue with the data rotation for two male NAmE speakers mentioned above, we applied another step of data normalisation to the articulatory data. For the lingual sensors, normalised values for each sensor were projected onto a common millimeter space. This was done by multiplying the *z*-scores by the mean standard deviation across

all sensors and then the overall mean was added. This allows us to present values in millimeters that retain the structure of the data. The same process was followed for the labial sensors. Thus, the millimeter values discussed in the context of individual differences are values that have been normalised in a manner comparable to our treatment of formants.

Chapter 3. Results

We report the acoustic results first, followed by the articulatory results, including both TD position and lip rounding, for both NAmE and AusE. Following the acoustic and articulatory overviews for each dialect we report correlations between acoustic and articulatory measurements of each vowel and dialect differences found in each type of data.

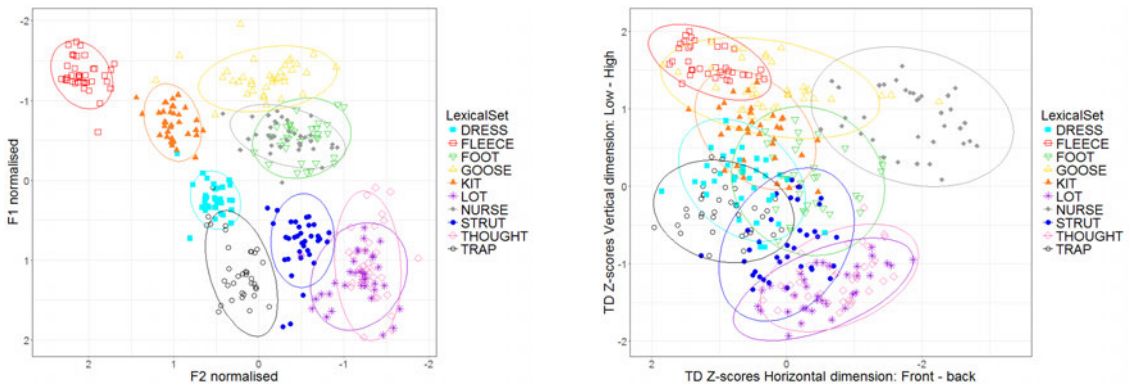
3.1. NAmE acoustics and articulation

3.1.1. Acoustic data overview

The distribution of normalised *F*1 and *F*2 values across the acoustic vowel space for NAmE is presented in Figure 3a. Ellipses represent 95% confidence intervals for each vowel, and are centred on the mean of each vowel category. Normalised *F*2 values are shown on the x-axis, and normalised *F*1 values are shown on the y-axis.

In the following discussion of the acoustic data, we refer to three groups of vowels – front, central, and back – based upon how the vowels are differentiated by relative *F*2 values. In grouping vowels based on *F*2, we consider the covariation of *F*2 and *F*1. Since *F*2 decreases with increases in *F*1, our groupings of front, central and back follow diagonals from the top left to the bottom right of the formant space. There are four vowels with comparatively high *F*2 that are differentiated by *F*1. In order of low to high *F*1, these

vowels are: FLEECE, KIT, DRESS and TRAP. We refer to these as front vowels. The front vowel with the lowest average *F*1, FLEECE, has an *F*2 that is nearly two standard deviations above the mean *F*2 while the front vowel with the highest average *F*1, TRAP, is near the mean value of *F*2 in the data. There are five vowels that have relatively low *F*2 values. We refer to these as back vowels and list them here in order from low to high *F*1: GOOSE, NURSE, FOOT, LOT, and THOUGHT. NURSE and FOOT are heavily overlapped in *F*1 and *F*2, but they are differentiated in *F*3. NURSE, which is rhotic for these speakers, has a lower mean *F*3 (*z*-score) than FOOT: mean *F*3 for NURSE = -2.181 (SD = 0.718), c.f. mean *F*3 for FOOT = -0.117 (SD = 0.371), a difference which is significant based on a linear mixed effects model[1] (β_{vowel} = -2.12, *SE* = 0.39, $t(4)$ = -5.41, p = 0.005), where SE is the Standard Error. The remaining vowel, STRUT, has an intermediate *F*2, which is lower than the front vowels, TRAP and DRESS, and higher than the back vowels, THOUGHT and LOT, of comparable *F*1. We refer to this vowel as central. We now turn to the articulatory data to observe how the differences in formant values correspond to tongue position in NAmE.



3a 3b

Figure 3. Normalised formants (a) and TD sensor positional coordinates (b) for NAmE vowels.

[1] Refer to linear mixed effects model explained in more detail in Section 3.4 (page 41).

3.1.2. Articulatory data overview

3.1.2.1. TD position.

In order to assess whether the vowels we have termed front, central, and back on the basis of formant measurements indeed correspond to front, central and back lingual articulatory positions, we first present data on the position of the TD sensor. The mapping from articulation to acoustics is of course impacted by differences in vocal tract area function across the entire length of the vocal tract. Focusing on a single fleshpoint necessarily has limitations but has frequently been used as a heuristic for tongue position in vowels (e.g., Noiray et al., 2014; Georgeton et al., 2014), and allows us to maintain comparability to past research. Besides the TD sensor we also explored the TB sensor and the point of inflection of a polynomial curve fit to the three lingual sensors. Of these measures, we found TD position to be the measure that best differentiated vowels within and across speakers.

Figure 3b shows the normalised values (*z*-scores) of the TD sensor for all five subjects. The y-axis shows the vertical position, and the x-axis shows the horizontal position from front (positive *z*-scores on the left side of the figure) to back (negative *z*-scores on the right side of the figure). As with the formant data, ellipses contain 95% confidence intervals for each vowel distribution and are centred on the mean. The distribution of the TD sensor follows the range of motion with which that fleshpoint on the tongue varies across vowels. Although there are some notable exceptions, by and large, vowels that are differentiated by *F*1 in the acoustics are differentiated by TD height. This is particularly clear for the front vowels. FLEECE, KIT, DRESS, and TRAP are all differentiated by tongue height, and the TD height differences are inversely related to *F*1. While we noted

covariation between *F*1 and *F*2 in acoustic space, we do not see corresponding covariation between the horizontal and vertical position of TD. For example, within the front vowels, FLEECE and TRAP are slightly more fronted than KIT and DRESS, c.f., the diagonal patterning of these vowels in the acoustic vowel space. On the basis of the formant data, we described NAmE as having one central vowel, STRUT. The TD data indicate that, in addition to STRUT, GOOSE and FOOT also have an intermediate degree of backness. The TD data indicate that GOOSE is more back than FLEECE, FOOT is more back than DRESS, and STRUT is more back than TRAP. The remaining vowels – NURSE, LOT, and THOUGHT – are even more back than GOOSE, FOOT, and STRUT. Of particular note is the fact that NURSE is produced with a considerably more retracted tongue position than FOOT, despite a similar *F*2 value.

3.1.2.2. Lip rounding.

Lip rounding involves protrusion of both the upper lip and the lower lip. Our metric of lip rounding is the average horizontal position of the UL and LL sensors. Figure 4 shows boxplots indicating the mean position in the *x*-dimension (horizontal) of the UL and LL sensors across vowels, normalised across speakers. These plots show that in our NAmE data, the most rounded vowel is GOOSE, followed by FOOT and NURSE. The notches in the boxplots indicate 95% confidence intervals around the median values. The confidence intervals for GOOSE, FOOT and NURSE do not overlap the other vowels, indicating statistically significant differences (at $\alpha = 0.05$), i.e., these three vowels are significantly more rounded than the other vowels. All else equal, an elongated vocal tract resulting from lip rounding is expected to lower formant values, particularly *F*2 (Stevens, 1989). As previously described, GOOSE, FOOT and NURSE are in the back vowel space based on

acoustic measures, which can be a consequence of different degrees of rounding and tongue backness.

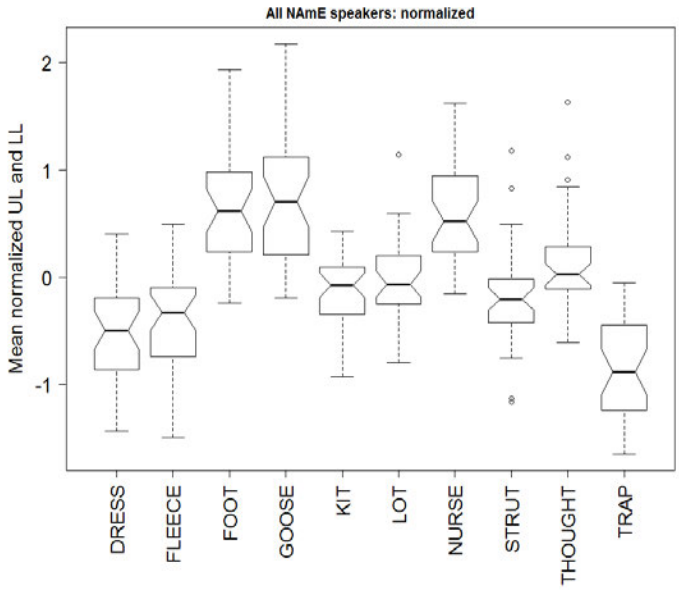


Figure 4. Box plots of the mean of the UL and LL position in the longitudinal dimensions (used as an index of lip rounding): NAmE. Notches indicate 95% Confidence Intervals around median values.

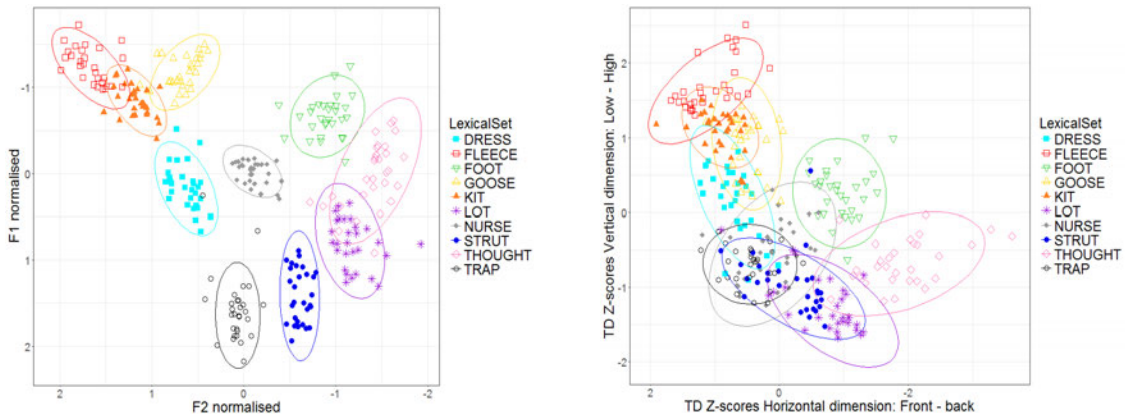
In summary, most of the NAmE vowels in this study can be clearly differentiated on the basis of *F*1 and *F*2. Exceptions to this are LOT and THOUGHT, which are overlapping, as well as NURSE and FOOT, which are distinguished by *F*3. The relative tongue positions for the vowels in NAmE show similar but not identical patterns. FLEECE, KIT, DRESS, and TRAP all have front TD positions. NURSE is the farthest back, probably because of its rhotic quality in NAmE. Of the non-rhotic vowels, GOOSE, FOOT, LOT and THOUGHT are produced with a more retracted tongue position than the other vowels; however, note that GOOSE has variable backness measurements. GOOSE, NURSE and FOOT were the most rounded of the vowels. Rounding for GOOSE and FOOT may contribute to an explanation of why these vowels show greater separation from front vowels FLEECE and KIT in *F*2 than they do in TD backness. LOT and THOUGHT are realised with the same TD

position, while both are further back than STRUT, which is central. Therefore, the acoustic vowel space for NAmE could be considered to display a 4:1:4/5 configuration, whereby there are four front vowels differing in height, one central vowel, and four or five back vowels, depending on whether LOT and THOUGHT are merged. However, based on tongue position alone, it appears the following might be a better description: 4:3:2/3 with FLEECE, KIT, DRESS and TRAP being front, NURSE, LOT and THOUGHT being back, and GOOSE, FOOT and STRUT being central. Thus the descriptions of the vowel space in terms of acoustics (*F*1 and *F*2) vs articulation (quantified as TD height and backness) lead to slightly different conclusions for the central and back vowels, which are not as clearly differentiated by TD position as are the front vowels.

3.2. AusE acoustics and articulation

3.2.1. Acoustic data overview

The distribution of normalised formant values (*F*1 and *F*2) across the acoustic vowel space is presented in Figure 5a. The ellipses show 95% confidence intervals for each vowel, and are centred on the mean of each vowel category (as for the NAmE data in Figure 3a). *F*1 and *F*2 are plotted on the y-axis and x-axis, respectively. In line with previous acoustic studies of AusE (e.g., Cox, 2006), the vowels are fairly evenly distributed across the vowel space and can be classified as front, central, and back on the basis of *F*2. There are four vowels with high *F*2, i.e., front vowels that differ in *F*1: FLEECE, KIT, DRESS, and TRAP. There are three central vowels that have intermediate *F*2 values, GOOSE, NURSE and STRUT, and also differ with respect to *F*1. The remaining back vowels have low *F*2: FOOT, THOUGHT, and LOT. We again turn to the articulatory data to observe how the differences in formant values correspond to tongue position in AusE.



5a 5b

Figure 5. Subfigures a and b display normalised formants and tongue dorsum sensor positional coordinates respectively for AusE vowels.

3.2.2. Articulatory data overview

3.2.2.1. TD position

Figure 5b shows the normalised values (*z*-scores) of the TD sensor for all four AusE subjects. The structure of the figure follows Figure 3b. The y-axis shows the vertical position, and the x-axis shows horizontal position from front (positive *z*-scores on the left side of the figure) to back (negative *z*-scores on the right side of the figure). Ellipses represent 95% confidence intervals and are centred on the mean position of each vowel. The distribution of vowels in the articulatory data generally follows the distribution of vowels in formant space, perhaps even more so than NAmE. More specifically, *F*1 tends to be inversely correlated with tongue height, and *F*2 tends to be inversely correlated with tongue backness. Of the front, central, and back vowels determined on the basis of the formants, the back vowels show the least overlap at the TD sensor. The back vowels – FOOT, THOUGHT, and LOT – are realised with a TD position that is more posterior than the other vowels. THOUGHT is the most back and FOOT and LOT are both realised with a

similar longitudinal position, with FOOT higher than LOT, as expected from the formant values. The centre of the ellipses for STRUT and NURSE are closest to zero on the x-axis, indicating that they are at the average level of backness in the data. We characterised these vowels, in addition to GOOSE, as central vowels by virtue of having intermediate *F*2 values. Of these three central vowels, GOOSE has the most front TD position. The front vowels FLEECE, KIT, DRESS, and TRAP have positions that are indeed more anterior than the other vowels.

3.2.2.2. Lip rounding

Figure 6 shows boxplots of the average horizontal position of the UL and LL sensors across vowels (c.f., NAmE in Figure 4). These data show that the most rounded vowels are FOOT, GOOSE and THOUGHT, possibly also NURSE, which is slightly different from our NAmE data. GOOSE is more rounded than the other central vowels NURSE and STRUT. Speaker M03 is the only speaker who deviates from this pattern. For him, THOUGHT is less rounded than for the other speakers such that THOUGHT shows a similar degree of rounding as LOT. We note that it is this speaker's tokens of THOUGHT that contributed to the overlap between THOUGHT and LOT ellipses in the vowel space in Figure 5a.

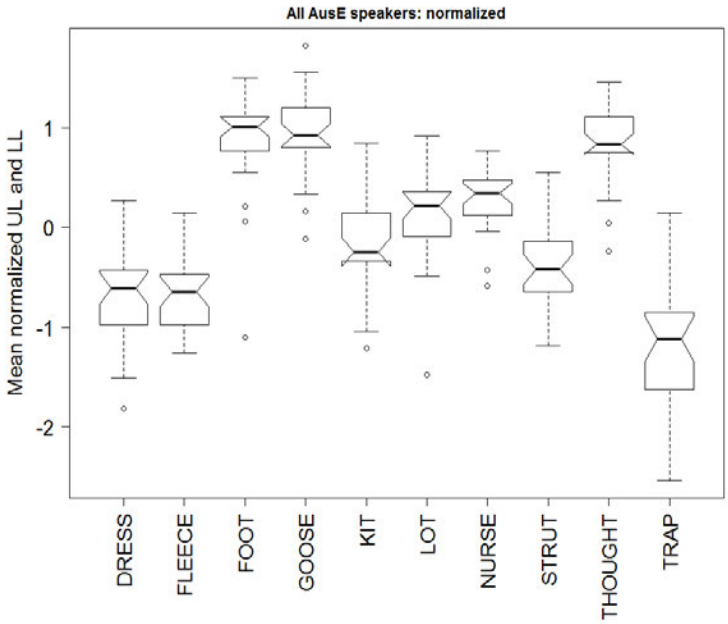


Figure 6. Box plots of the mean of the UL and LL data in the longitudinal dimension (used as an index of lip rounding): AusE. Notches indicate 95% Confidence Intervals around median values.

In summary, the relative tongue positions for the vowels in AusE are generally as expected from the formant values, given the common heuristics deployed in dialect comparison. By and large, the AusE vowels in this study can be differentiated on the basis of *F*1 and *F*2, and a similar partitioning of the vowel space can be observed in the position of the TD sensor in vertical and longitudinal dimensions, although with greater overlap. The AusE vowel space can be considered to display a 4:3:3 configuration, whereby there are four front vowels differing in height, three central vowels differing in height, and "three" back vowels also differing in height. In the following section we examine the correlations between acoustic and articulatory data.

3.3. Acoustic-articulatory relations

In the two previous sections we provided a general overview of the acoustics and articulation of both NAmE and AusE based on nine speakers in total. In this section we

examine the acoustic-articulatory relation more directly. We evaluate linear relations between formant values and TD position across dialects and within dialects and also uncover cases in which the linear relation breaks down.

Pearson product-moment correlations were computed to quantify the relationship between formants and tongue position. Across dialects there was a strong negative correlation between *F*1 and TD height, $r = -0.78$, $p < 0.001$, and a positive correlation between *F*2 and TD backness, $r = 0.69$, $p < 0.001$. These correlations are in the expected directions, since, in our data, the vertical coordinate increases with TD height while the longitudinal coordinate decreases with TD backness. Correlations within dialect produce similar results, slightly stronger negative *F*1/TD height correlations and moderate to strong positive correlations for *F*2/TD backness: NAmE, *F*1/TD height, $r = -0.817$, $p < 0.001$ and for *F*2/TD backness, $r = 0.563$, $p < 0.001$; AusE, *F*1/TD height, $r = -0.741$, $p < 0.001$ and *F*2/TD backness, $r = 0.811$, $p < 0.001$.

Although there are reasonably strong correlations both within and across dialects, we also noticed that linear correlations were stronger for some speakers than for others. For one AusE male speaker (M03, see Figures 7 a and b), we observed an acoustic-articulatory mismatch in backness for vowels LOT and THOUGHT. The TD is further back for THOUGHT than for LOT (Figure 7b) but THOUGHT has a higher *F*2 than LOT (Figure 7a). In this case, *F*2 provides a poor diagnostic for tongue backness. The lip data revealed that this speaker produced a smaller difference in rounding for this vowel pair, which offers a likely explanation for why this speaker showed a similar lingual articulatory pattern but a different *F*2 pattern from the other AusE speakers. It is possible that for these vowels lip rounding has more of an impact on *F*2 than tongue backness.

Another mismatch in the data is less easy to explain. For one of the AusE male speakers, M05, we observed an inconsistency in the acoustic-articulatory relation in the central part of the vowel space (Figures 7c and 7d). Although this speaker shows the same degree of acoustic-articulatory correspondence as other speakers in the front and back section of the vowel space, the central vowels NURSE, GOOSE and STRUT all have a similar level of backness i.e., as determined by the horizontal position of TD, while displaying large differences in *F*2.

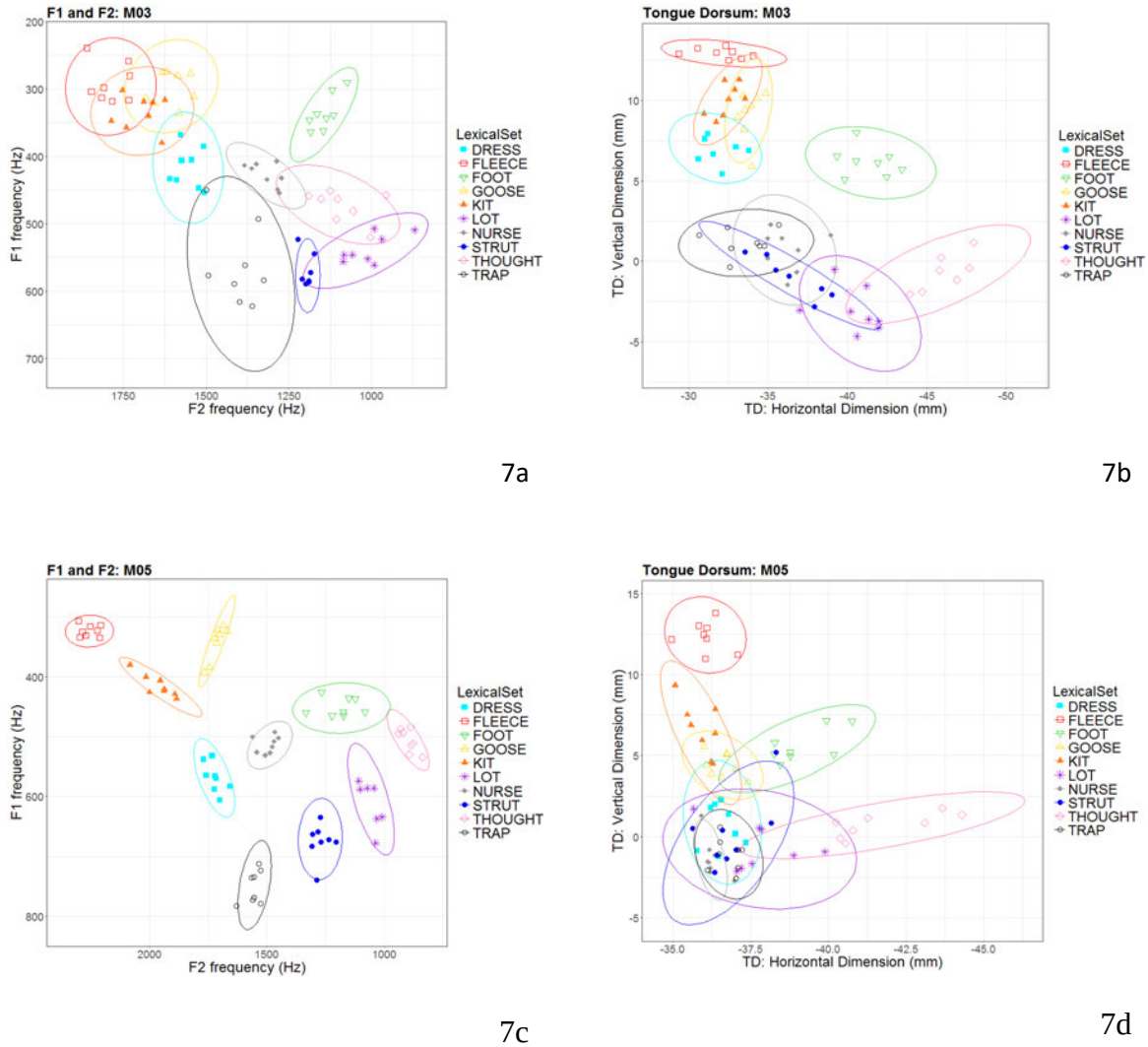
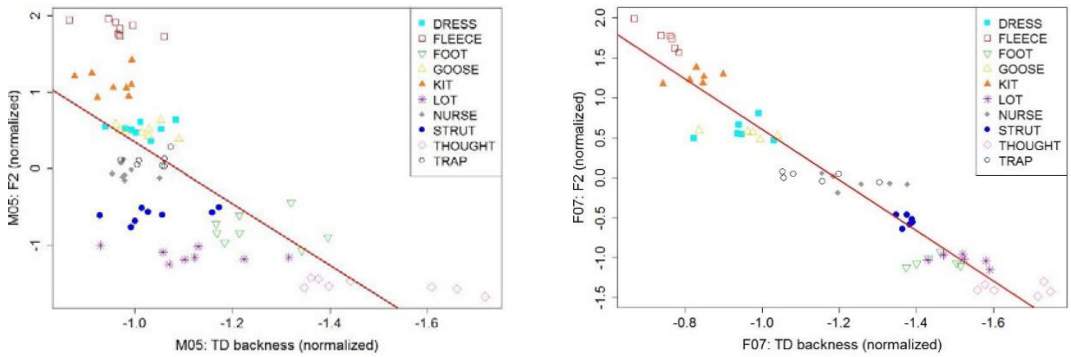


Figure 7. (a) and (c) on the left display individual speaker F1 and F2 values. (b) and (d) on the right display individual speaker TD sensor positional coordinates for the speaker on the same row.

Unlike the case of M03's THOUGHT and LOT vowels described above, it is unlikely that the difference in M05's *F*2 across GOOSE, NURSE, and STRUT is due to a degree difference in lip rounding. This speaker follows the AusE group trend for lip rounding; GOOSE is the most rounded vowel followed by NURSE. However, for this speaker, it is GOOSE that shows unexpectedly high *F*2 values given the TD position. Rounding would be expected to lower *F*2, the opposite pattern of what we observed. The relationship between *F*2 and tongue backness for M05 can be seen in Figure 8a. For reference, we have plotted the same

relation for another AusE speaker in Figure 8b. Vowels are differentiated by colour and symbol, with a regression line fit to the data points. For speaker M05, tokens of GOOSE appear above the regression line (for *F*2-TD backness) while tokens of NURSE fall below it. For F07, the relationship between *F*2 and TD backness is more linear, and is in line with expectations based on the acoustic-articulatory relations data presented above. However for M05 there appears to be a non-linear relationship for these aspects of acoustics and articulation. Thus, although we find strong correlations between formant values and TD position, there are also corners of the data in which such correspondences break down.



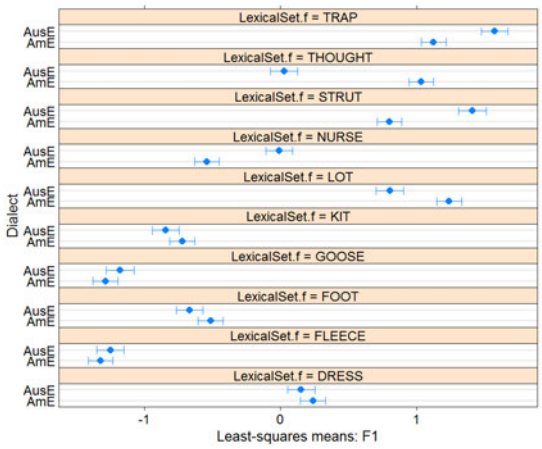
8a 8b

Figure 8. (a) and (b) display normalised F2 plotted against TD backness for AusE speakers F07 and M05, respectively.

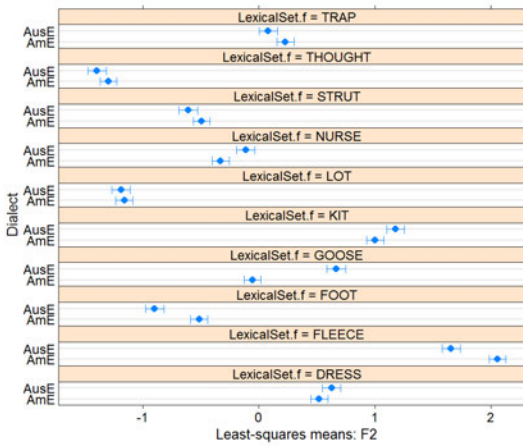
3.4. Dialect comparison

To quantify the difference between the two dialects we fit linear mixed-effects models to *F*1, TD height, *F*2, and TD backness. The fixed factors in the models were lexical set (i.e., vowel) and dialect as well as the interaction between these factors. Random variables

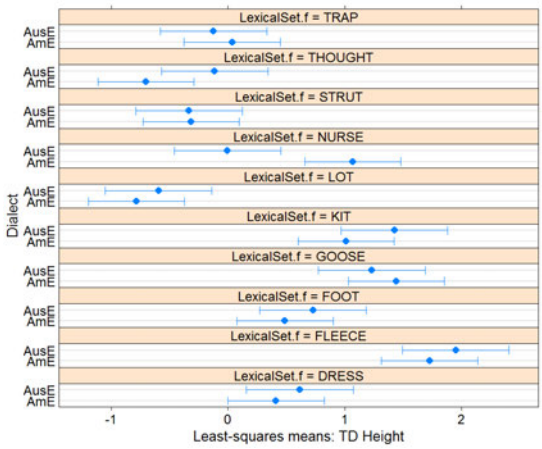
were speaker and block presentation order, whether the vowels were produced in one of the blocks early in the experiment or in one of the blocks later in the experiment. Tables summarising the models can be found in Appendix 1. The residuals of all models were checked for normality and heteroscedasticity. Dialect was not a significant predictor but the interactions between lexical set and dialect were significant for all dependent measures. To visualise these results, we plotted model predictions for the interaction term. Figure 9 shows the estimated marginal means, or predicted means, with 95% confidence intervals for each dialect by lexical set (Figure 9). Figure 9a shows the results for *F*1 (where *F*1 increases from left to right) and Figure 9b shows *F*2 (where *F*2 increases from left to right). Figure 9c shows the results for TD height (where tongue height increases from left to right) and TD backness is represented in Figure 9d (where tongue backness decreases from left to right, i.e., the tongue is in its most anterior position at the far right).



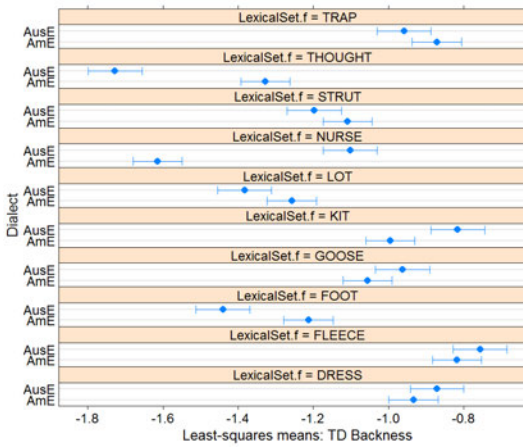
9a



9b



9c



9d

Figure 9. (a-b) display least square means (estimated marginal means) for dialect by vowel with 95% Confidence Intervals for F1 and F2 respectively; (c-d) display least square means as above for TDz and TDx respectively (the x-axis from left to right represents TDz positions increasing in height; the x-axis from left to right represents TDx positions decreasing in backness). All measurements normalised across both dialects.

All vowels except for DRESS differ significantly across dialects in either *F*1 or *F*2 (or both *F*1 and *F*2). TD position differentiates fewer vowels. In general, a difference in TD position across dialects implies a difference in formants—there is just one exception. However, a significant difference in formants does not imply a significant difference in TD position. Five vowels differ across dialects in *F*1 (TRAP, THOUGHT, STRUT, NURSE, LOT),

but only one of these differs in TD height (NURSE). *F*1 was higher for NAmE THOUGHT and LOT than for AusE THOUGHT and LOT. These differences in *F*1 do not correspond to significant differences in TD height. Particularly for LOT, the dialects are very similar in TD height despite the significant *F*1 differences. The only vowel with a significant difference across dialects in TD height was NURSE. This difference has the expected corresponding difference in *F*1, i.e., NAmE differs from AusE in having both lower *F*1 and higher TD position. Six vowels differ across dialects in *F*2 (TRAP, NURSE, KIT, GOOSE, FOOT, FLEECE). Three of these show corresponding significant differences in TD backness (NURSE, KIT, FOOT). Finally, there was one significant difference in TD backness that did not have a corresponding significant difference in *F*2—this was for the vowel THOUGHT, which has a considerably more anterior TD position in NAmE than in AusE.

Chapter 4. Discussion

4.1. Correspondences between acoustics and articulation

Parallel acoustic and articulatory data on monophthongs from two English dialects, NAmE and AusE, have allowed us to evaluate the correspondence between articulation and acoustics that is frequently used to reason about how variation in formants across dialects relates to articulation. Variation in *F*1 is assumed to correlate inversely with tongue height; while variation in *F*2 is assumed to correlate with tongue backness.

As was introduced at the beginning of the thesis, these heuristics relating *F*1 and *F*2 to TD position have theoretical bases in tube models of the vocal tract which predict this correlation only within certain ranges of articulatory variation (e.g., Chiba & Yokoyama, 1941; Fant, 1960; Stevens, 1989). For example, Stevens (1989) suggests a two tube model for low vowels, e.g., STRUT, LOT, THOUGHT, and a three-tube model with a Helmholtz resonator for vowels with a narrow constriction, e.g., GOOSE. Given the boundary conditions for the Helmholtz resonance defined for the three-tube model, *F*1 is predicted to increase as the cross-sectional area of the constriction widens. This should give rise to a linear (or quasi-linear) correlation between *F*1 and TD height. Moreover, when the area of the constriction widens beyond the range of values that can support Helmholtz resonance, a further increase in *F*1 is expected in the transition from a three-tube to a two-tube model. This would also contribute to the correlation between *F*1 and TD height.

These two mechanisms, increasing the cross-sectional area of the constriction supporting Helmholtz resonance and transitions from vocal tract shapes that support Helmholtz resonance to those that do not, both conspire to yield correlations between *F*1 and TD height. There are also conditions expected to give rise to correlations between *F*2 and TD backness. The nomograms of Stevens (1989) show that advancement of a vocal tract constriction will raise *F*2, if the constriction is in the posterior part of the vocal tract for a three-tube model (high vowels) or the anterior portion of the vocal tract for a two-tube model (low vowels). Outside of these regions, advancement of TD position can have minimal effect on *F*2 or even lower *F*2, e.g., anterior constrictions in a three-tube model or posterior constrictions in a two-tube model. From this theoretical standpoint, the stronger correlations between *F*1 and TD height than for *F*2 and TD backness in our study are not particularly surprising. More importantly, we can predict the conditions under which the simple heuristic will break down.

By and large, data from a single fleshpoint on the tongue dorsum displayed articulatory patterning across vowels that correspond to those in the formants. In particular, the relative lingual height and backness of vowels at the TD sensor correlates with *F*1 and *F*2 values. Wieling et al. (2016) is one of the few studies that reports correlations between EMA data and formant values which can serve as a basis for comparing the strength of the correlations in our data. However, for other point-tracking studies, as mentioned in Chapter 1, see also Iskarous (2010), Dromey, Jang, and Hollis (2013), Whalen, Chen, Tiede and Nam (2018), and Gorman and Kirkham (2020) for EMA studies and Iskarous (2001 & 2010), and McGowan and Berger (2009) for additional studies looking at correlations between formant and articulatory data in the X-Ray

MicroBeam corpus. In Wieling et al.'s (2016) corpus of Dutch speakers from Ter Appel and Ubbergen, they report correlations between *F*1 and tongue height of r = -0.22 for one set of words and r = -0.43 for another. These correlations are much weaker than the r = -0.78 correlation found in our data. It is not clear what causes this discrepancy across studies, but there are several methodological differences that may play a role. Our correlations include just one pair of articulatory and acoustic data points per token. The measurements were made at the point of minimum velocity in the movement, a proxy for the target of controlled movement. Wieling et al. (2016) included multiple such pairings per token in their correlations. Accordingly, the correlations represent both within-token and across-token variability. Another difference is that our vowels were produced in a consistent phonetic frame whereas the vowels in Wieling et al. (2016) came from a diverse range of phonetic environments. These methodological differences could have reduced the consistency with which a single fleshpoint is predictive of *F*1. There may also be real linguistic differences across Dutch and English that contribute to how well TD height corresponds to *F*1. The correlations between *F*2 and tongue backness across studies were more comparable. Wieling et al. (2016) report r = -0.44 for one set of words and r = -0.63 for another (c.f., r = 0.69 in the current study). Due in part to the weak correlations, Wieling et al. cautions about interpreting *F*1 and *F*2 in terms of tongue position. We concur with this precaution, but we also seek to understand the conditions under which formant values are more or less predictive of tongue dorsum position.

For starters, we found correlations between *F*1 and *F*2 in the front part of the vowel space that do not correspond closely to the vertical and longitudinal displacement of the tongue dorsum. Differences in *F*2 amongst the front vowels within both English dialects

investigated in the current study were a result of general properties of formant spaces: as *F*1 increases, *F*2 of front vowels also tends to decrease, leading to a diagonal distribution on the vowel quadrilateral. We assume that these differences in *F*2 amongst the front vowels are at least in part attributable to an intrinsic relationship between tongue height and pharyngeal area due to the conservation of tongue volume and, thus, may not be under speaker control to the same degree as *F*2 in other locations of the vowel space. Decreases in *F*2 as a function of *F*1 in the front part of the vowel space were consistent across dialects and speakers, regardless of TD backness.

As another general observation, the vowel space expressed in terms of TD position is more compact than the vowel space expressed in formants in that there was more overlap in the TD positional coordinates between some vowels than we observed in the formant plots.[2] From this we surmise that other aspects of vowel articulation function to modulate the impact that TD position has on vocal tract resonance. In some cases, we observed that vowels with similar tongue positions, e.g., KIT and GOOSE in AusE, are differentiated by lip rounding. Incorporating other aspects of articulation – e.g., jaw height (Erickson, 2002; Stone & Vatikiotis-Bateson, 1995), tongue shape (Dawson et al., 2016), or additional data points on the tongue may provide a more dispersed view of the articulatory vowel space, and a closer correspondence to the acoustics. In what follows we consider some of these factors in the context of a broader discussion of the degree to which tongue position can be inferred from vowel formants.

[2] It must be noted here that this does not necessarily mean that there was less variability in the articulation for each vowel as compared to the formants.

Although linear correlations between TD position and formant values were stronger in our data than other similar studies, we observed individual differences in the strength of correlations. In particular, we described some discrepancies in the expected acoustic-articulatory relations for two of the male AusE speakers, M03 and M05. In the case of M03, lingual articulation for THOUGHT and LOT was similar to the other AusE speakers with THOUGHT more retracted than LOT, but his formant values showed higher *F*2 for THOUGHT than LOT. This difference is consistent with the differences in lip rounding that we also observed. Unlike other AusE speakers, this speaker did not differentiate THOUGHT and LOT in rounding.

Another AusE speaker, M05, showed particularly weak linear correlations between *F*2 and TD backness. The *F*2 values for M05 tend to be above the regression line at high and low values of tongue backness and below the regression line at intermediate values, indicating a non-linear trend, which contrasts to the largely linear trend observed for other speakers (e.g., F07 in Figure 9). Given the particular anatomy of M05, central vowels may fall into an area of stability such that variation in TD backness has little effect on *F*2. An alternative hypothesis is that something else is influencing *F*2 other than TD backness. One possibility may be tongue curvature. Tongue shape has been shown to differentiate the vowels of English (Dawson et al., 2016). A more curved tongue would lead to a larger pharyngeal cavity which in turn would result in an increase in *F*2, due to an increase in the cross-sectional area of the back cavity (while holding constriction location constant). Further investigation would be needed to discover why *F*2 varies despite similar degrees of TD backness for these central vowels, and why this is the case in this part of the vowel space and for this speaker in particular.

Cases such as these underscore the indeterminacy of interpreting formant values in terms of lingual articulation, or at least with regard to a single fleshpoint. Because they are shaped by multiple articulatory constrictions in the vocal tract, it is not always possible to map changes in formants to changes in TD position. Gorman and Kirkham (2020) also make the point that phenomena like small sublingual cavities cannot be adequately captured by EMA point tracking and that this results in unmeasured aspects of vocal tract shaping which have potential to influence acoustic output, producing apparent acoustic-articulatory mismatch.

4.2. Differences between dialects

Turning now to a comparison between dialects, we discover that whether acoustic or articulatory data are examined changes our conclusions about how vowels differ across NAmE and AusE. Differences between the dialects uncovered in this study using both methods are discussed below.

The acoustic results we reported for NAmE and AusE largely replicated past acoustic studies on these dialects. In terms of formants, all vowels except DRESS differ significantly across dialects in either *F*1 or *F*2. The front vowels were similar across dialects (although note that TRAP differs significantly on *F*1 and FLEECE differs significantly on *F*2). There were several differences in the central and back vowels between dialects. NAmE can be characterised acoustically (on the basis of *F*2) as having just one central vowel, STRUT, while AusE has three, STRUT, NURSE, and GOOSE. In the back vowels, our NAmE speakers

tended to merge THOUGHT and LOT. Three speakers made no distinction and two showed overlapping distributions. All AusE speakers, on the other hand, maintained a clear distinction at least in *F*1. Three out of four AusE speakers also differentiated THOUGHT and LOT in *F*2, with THOUGHT having a lower *F*2 than LOT (we discussed lip rounding as the basis for this exception above). Thus, on the basis of the acoustic results, we partitioned the AusE vowel space into four front, three central, and three back vowels, or 4:3:3 and the NAmE vowel space into 4:1:4/5. Both dialects share STRUT as central vowel but differ in whether the other non-front vowels are central or back.

As raised in the introduction, vowel spaces based on formants are sometimes assumed to have clear lingual articulatory correlates. We have found that the articulatory data reveals a different partition of the NAmE vowel space, which has implications for how the differences between dialects are characterised. The key difference between the acoustic and articulatory characterisation of NAmE vowels involved the position of GOOSE and FOOT, which have low *F*2 values but central TD position. These are both rounded vowels (Figure 4), and the rounding no doubt lowers *F*2 beyond what would be expected from TD position alone. In the absence of the comparison with AusE, we might even be tempted to conclude that lip rounding is the source of the discrepancy between formant values and TD position for these vowels in NAmE. Viewed in light of the comparison with AusE, this conclusion appears incomplete. GOOSE and FOOT are also rounded in AusE, and to a similar degree as in NAmE (Figure 6). Despite similar degrees of rounding, there are significant differences across dialects. FOOT is a vowel that differs significantly in both *F*2 and TD backness—it is both further back and has a lower *F*2 in AusE than in NAmE. The comparison with AusE makes it clear that the TD position for

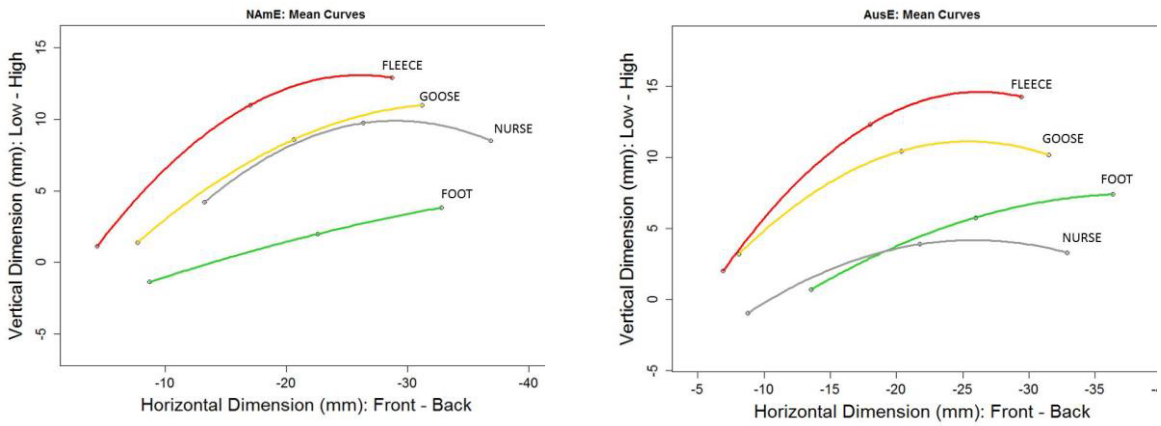
FOOT in NAmE is more anterior, even though *F*2 is still relatively low. The same goes for GOOSE. It has a central TD position in NAmE. Across dialects, GOOSE showed significant differences in *F*2 (AusE GOOSE is higher in *F*2), without a corresponding difference in TD backness. Like FOOT, *F*2 for NAmE GOOSE is low despite an advanced (central) TD position. Thus, from an articulatory perspective, both dialects have GOOSE as a central vowel, rather than back. In the formant space, however, GOOSE is back in NAmE and central in AusE. In this case, how to partition the vowels into front, central and back, depends on whether we refer to the vowel space based on TD position or the vowel space based on formants.

The case of GOOSE-fronting in NAmE without a corresponding rise in *F*2 highlights the need to incorporate articulatory parameters besides TD position and rounding into our understanding of formant variation and our description of dialects. This phenomenon of GOOSE-fronting was outlined in the introduction. Recall that in studies of English dialect variation, "GOOSE-fronting" refers to an increase of *F*2 in the GOOSE vowel such that the GOOSE category encroaches on FLEECE and KIT. In some dialects of English, GOOSE-fronting was initiated by high frequency words in which GOOSE is followed by a coronal stop (e.g., Derby English; Sóskuthy et al., 2015). Given this environment, GOOSE-fronting is thought to be driven by a coarticulatory effect of the tongue being pulled forward during the vowel in anticipation of the coronal articulation (Harrington et al., 2008). Viewing the acoustic data together with the articulatory data presents a more nuanced view. As expected, GOOSE is more front in the acoustic data for AusE than NAmE, in that the *F*2 value for GOOSE is closer to the *F*2 values of FLEECE and KIT in AusE than in NAmE. The prediction made in the introduction was that the tongue would be more

anterior for GOOSE in AusE speakers compared to NAmE speakers. However, in articulation, NAmE GOOSE is also "fronted", i.e., not significantly different from AusE.

To better understand how NAmE GOOSE could be fronted articulatorily without raising *F*2, we explored tongue curvature, which we computed by fitting a second-order polynomial to the three sensors on the tongue. Figure 10 shows the mean positions of the lingual sensors with the fitted polynomial for a subset of vowels: NURSE, FOOT, GOOSE, and FLEECE. Figure 10a (left) shows the NAmE data; Figure 10b (right) shows AusE. As illustrated in Figure 10a, FLEECE is more curved than GOOSE in NAmE. The mean quadratic term for NAmE FLEECE is -1.54(SD = 0.45) and the mean quadratic term for GOOSE = -0.46(SD = 0.43)). In AusE (Figure 10b), the difference in curvature between FLEECE and GOOSE is not as large. The mean quadratic term for AusE FLEECE is -1.6(SD = 0.35), c.f., mean quadratic term for GOOSE = -1.01(SD = 0.39)). We confirmed the statistical significance of dialect differences in curvature by fitting a mixed effects model with dialect as a fixed factor (and speaker and order as random effects) to the quadratic term from the polynomial function for GOOSE tokens and for FLEECE tokens. AusE GOOSE was significantly more curved than NAmE GOOSE (β = 0.965, *SE* = 0.224, $t(7)$ = 4.303, p = 0.0035) but the effect of dialect on FLEECE was not significant (i.e., FLEECE was not significantly more curved in one dialect than the other). Based on this result, it is tempting to conclude that the differences in *F*2 for GOOSE across dialects are attributable (at least in part) to differences in tongue curvature, at least for the subjects reported here. A more curved tongue may be indicative of a larger pharyngeal cavity, which would have the effect of increasing *F*2. We note in this context that other factors, including palate shape, may also influence tongue curvature (Lammert et al., 2013). Although we cannot provide

conclusive evidence, the larger *F*2 difference between GOOSE and FLEECE in NAmE than in AusE may be a consequence of tongue shape, rather than tongue position.



10a 10b

Figure 10. (a) displays mean tongue curves for four NAmE vowels: FLEECE, GOOSE, NURSE and FOOT. (b) displays tongue curves for the corresponding AusE vowels. The three circles on each curve represent the three lingual sensors (from left to right: TT, TB, and TD).

The vowels NURSE and FOOT offer additional cases in which curvature appears to be modulating the relation between formant values and TD position. These vowels are not distinguished acoustically in *F*1 and *F*2 for NAmE. Figure 3a shows the degree of overlap for NURSE and FOOT in NAmE in the formant space and Figures 10a and 10b show the tongue shape for NURSE to be more curved than for FOOT, with a higher tongue that is also more back, which is the general pattern for NAmE. In contrast, these two vowels are clearly differentiated acoustically in AusE. From the formant plots alone, we might conclude that these vowels are similar in NAmE but different in AusE. The articulatory data reveal clear differences between the vowels in both dialects. The reason why they are merged in NAmE *F*1/*F*2 formant space is likely that being high and curved (NURSE) offsets the effect of TD backness on *F*2. Thus, NURSE is further back than FOOT even though these vowels have similar *F*2. Given the lowered *F*3 found for NURSE, TD

retraction likely corresponds to a constriction at the soft palate and/or pharynx (but see Espy-Wilson et al., 2000 for claims that the pharyngeal constriction is not responsible for lowering *F*3 in American English /ɹ/).

More broadly, we can see that curvature is an articulatory parameter on which vowels and dialects differ. In NAmE, FLEECE and NURSE are curved; GOOSE and FOOT are not. In AusE, FLEECE, GOOSE, and NURSE pattern together as curved to the exclusion of FOOT. Incorporating curvature into our description allows us to observe a similarity across dialects in the NURSE vowel that we would have missed otherwise. NURSE was the only vowel that was significantly different on *F*1, *F*2, TD height and TD backness, all four of the dependent variables reported in Figure 9. Despite these numerous differences as well as a difference in rhoticity, NURSE is curved in both dialects.

As reported in the introduction, past research had identified differences between GOOSE and NURSE vowels across dialects. Both GOOSE and NURSE have a higher *F*2 in AusE than in NAmE. According to our results, it is clear that the increase in *F*2 in AusE compared to NAmE has a different articulatory basis for the two vowels. The articulatory basis of *F*2 differences for NURSE are tongue height and backness, as this vowel is curved in both dialects. For GOOSE, the difference in *F*2 must be attributed to some other articulatory property, which we have suggested is tongue curvature.

Although there were other vowels besides GOOSE for which significant differences in formants do not correspond to significant differences in TD position, the reverse was rare. A significant difference in TD position typically implied a significant difference in formant values. The one exception to this trend was the THOUGHT vowel. For this vowel, there

was a significant difference in TD backness (AusE is further back than NAmE) but no difference in *F*2. Moreover, THOUGHT is rounded in AusE but not NAmE, which should, if anything, further increase the difference in *F*2 expected on the basis of tongue backness. Thus, the THOUGHT vowel is a clear case in which the simple heuristic relating *F*2 to tongue backness breaks down. Assuming that the vocal tract is partitioned into two cavities for THOUGHT with the partition leaving the front cavity larger than the back cavity, the heuristic fails because the theoretical basis for the *F*2 by TD backness correlation is not valid for this configuration. The TD position for THOUGHT in both dialects may fall within a region of stability within which variation in articulatory position exacts little influence on *F*2 (Stevens, 1989).

More likely, however, the constriction in AusE is posterior to this quantal region. The gradual advancement of a relatively posterior constriction in a two-tube model is predicted to lower (not raise) both *F*2 and *F*1. We observe the predicted effect on *F*1. *F*1 is lower for AusE, which has the more retracted TD position. We do not observe *F*2 differences for THOUGHT across dialects, but this may be because the influence of TD backness on *F*2 is offset by lip rounding. Overall, then, while THOUGHT defies the simple heuristic relating TD backness to *F*2, the articulatory-acoustic relation for this vowel and the differences across dialects are well-behaved from the theoretical foundations from which the simple heuristic was formulated.

4.3. More recent research

Since the current experimental work was conducted, research has been published that evaluates assumptions regarding acoustic-articulatory relations, and comparing

dialects on the basis of both articulation and acoustics. This includes papers in *The Journal of the Acoustical Society of America* special issue in which this research appeared, pointing out where the heuristics break down (e.g. see Lee et al., 2016; Strycharczuk & Scobbie, 2017; Wieling & Tiede, 2017). Other more recent research has commented directly on the findings reported here. For example, the work has contributed to a recent detailed description of the phonetics of AusE (Cox, 2019). Also, Gorman and Kirkham (2020) justified their research in part based on the current experiment's finding of a non-linear relationship between *F*2 and tongue backing for some vowels. The authors point to this experiment as an example of work that has "uncovered varying degrees of acoustic-articulatory mismatch in even relatively well-understood phenomena" (page 724).

Interestingly, two papers both citing this study also showed acoustic-articulatory mismatches in different English dialects of Britain, but they drew different conclusions regarding the reason for these differences. One suggested that dialect change may be detected in the articulation before it is realised in the acoustics (Gorman & Kirkham, 2020), and the other that a difference in systematic articulatory strategies for GOOSE fronting was a result of diachronic dialect change (Lawson et al., 2019). Using ultrasound, Lawson et al., (2019) looked at systematic differences in production strategies (or what they call "performative variation") for GOOSE in English dialects from England, Ireland and Scotland. They found that the Scottish English dialect employed the strategy of a more back tongue position with minimal lip protrusion compared to the other dialects which used a high-front tongue position supported by lip protrusion. They argued that these different strategies were not due to trading relations or motor equivalence (see Perkell et al., 1993), but rather due to a shift to a fronted GOOSE occurring diachronically (indeed,

centuries apart). Thus they support the findings here of an acoustic-articulatory mismatch for GOOSE whereby complementary strategies were used, however they argue against them being motor equivalence strategies *per se*. Gorman and Kirkham (2020) found unexpected *F*2 raising with TD advancement coupled with decreased lip protrusion over time for FOOT. They argue that in progress sound changes may involve speakers slightly modifying vocal tract articulations and that these would then take time to be realised in a stable acoustic-articulatory relationship. They would need to first establish a 'position' in a quantal part of the vocal tract. These differing conclusions of the above studies prompt speculation regarding the findings of this thesis. Could the increase in tongue curvature for GOOSE in NAmE be some indicator or precursor to a fronting phenomenon?

Many studies investigating both the acoustics and articulation of a dialect limit their analyses to a small number of vowels, however, as I have hopefully shown here, it can also be useful to take a broader brush stroke approach, assessing the majority of the vowels of a dialect, (see also Lee et al., 2016; Whalen et al., 2018; Wieling et al., 2016).

4.4. Study limitations

It is important to acknowledge some limitations of the present study before presenting ideas about future research and final conclusions.

The number of speakers (n = 9) was one limiting factor which could be remedied in future studies. However, reported multiple tokens for each of 10 monophthongs make this the largest study of parallel acoustic and articulatory data of AusE. Our four AusE speakers were from the same region, but the five NAmE speakers recorded for

comparison came from different parts of North America. Overall, the NAmE data showed more variation across speakers than the AusE data, which is likely due as least in part to the regional heterogeneity of the NAmE group. It is also possible that the NAmE group reported here may have been influenced by their time in Australia (Campbell-Kibler et al., 2014). Including speakers resident in North America might be a more reliable way of uncovering dialectal differences. However, we also note that some vowels may remain stable even after moving countries. For example, Nycz (2013) reported stability in low back vowel realisation for mobile Canadians. Nevertheless, at least one of our NAmE speakers showed signs of adopting AusE vowels in both acoustic and articulatory data, so including speakers resident in North America (or more recently arrived in Australia) with less regional variation might have resulted in clearer differences between the two groups than were reported here. The speaker variation for NAmE may also have strengthened some of the correlations we reported. From the standpoint of assessing the acoustics-articulation relation, variation in articulation provides a way to "sample" the space of possible articulations while observing the acoustic consequences. Perhaps another important factor to control with subjects is gender. If a larger sample size were possible, analysing male and female speech data would be of interest due to their vocal tract size differences.

EMA as a tool for investigating articulation has many advantages, as outlined in Chapter 2, however it is also not without its limitations (Kochetov, 2020). With EMA, the lingual data extracted is reduced to a small number of points, in this case three, and the tongue back area is not accessible due to the gag reflex. Although tongue curvature can be extrapolated by fitting splines to the curve as done in this study, information about the

overall vocal tract shape is missing, and particularly useful would be to have some measure for the pharyngeal cavity.

As mentioned earlier in the discussion, this study found greater correlations between acoustic and articulatory measures than Wieling et al., (2016). This was attributed in part to the vowel context and connected speech, and partly to the dynamic measures taken. For the purposes of this study, I believe the static measures were appropriate for identifying the vowel target, given monophthongs were being studied. Perhaps introducing some more variation by varying the context might have enabled a better representation of the vowel target in both the acoustic and articulatory data.

These limitations notwithstanding, the comparison offered in the current study presents an informative case study both of how articulatory data can enhance dialect description but also of how dialect variation can provide an informative domain for advancing understanding of the acoustics-articulation relation in speech.

4.5. Future research

A reflection follows on the direction future research might take, and then a brief presentation of some preliminary investigations. As discussed in this thesis, existing acoustic models predict ranges of vocal tract shapes over which the following acoustic-articulatory relations generally hold: $F1$ is inversely correlated with tongue height, and $F2$ is inversely correlated with tongue backness. For example, according to two-tube models (relevant for non-high vowels), variation in constriction location within the anterior portion of the vocal tract establishes boundary conditions for $F2$ (Stevens, 1989).

The evidence presented in this thesis indicates that the degree to which *F*2 and TD position are linearly correlated varies across speakers; this evidence comes from the strength of correlations computed across monophthongs. The results presented in the thesis for static (single time point) measurements of monophthongs make predictions for how the relations between *F*2 and TD position will vary for diphthongs, i.e., vowels that evolve dynamically from one vocal tract shape to another. In what follows, I outline details and preliminary results for such a study. This section is included to show the potential for these data to shed light on the acoustic-articulatory relationship, in particular, how well TD position might predict *F*1 and *F*2 throughout the trajectory of a diphthong.

4.5.1. Preliminary investigation: Aims and predictions

Dynamic measurements of diphthongs allow further tests of how the acoustic-articulatory relation varies as a function of vocal tract shape. The aim of this preliminary investigation, is to add articulatory and acoustic descriptions of diphthongs to further investigate the acoustic-articulatory relationship. Based on the monophthong results, it is possible to make predictions such as under what conditions the linear relation between *F*2 and TDx would be stronger or weaker. So, for a diphthong where the tongue moves through a trajectory (through space) it might be possible to capture this changing relationship throughout the vowel. Thus, it is possible to make specific predictions for vowels as to how the articulation affects the acoustics when the tongue moves from one configuration (e.g. low and back) to a different configuration (e.g. high and front, such as in '*soid*' (the CHOICE vowel) throughout the time interval of the vowel. The prediction, based on a two-tube model for non-high vowels would be that TDx advancement in the posterior region of the vocal tract should lower *F*2 while TDx advancement in more

anterior regions should raise *F*2. Therefore, TDx should predict *F*2 less well the higher the value of the TDx, i.e. for more back positions.

4.5.2. Preliminary investigation: Method

The same corpus of AusE speakers was used for this analysis. The AusE corpus included five diphthongs, also in the sVd context: FACE, PRICE, CHOICE, GOAT and MOUTH, as per Wells' (1982) lexical sets. Analysis of the diphthongs using dynamic measurements is presented here for three speakers: formants and lingual (TD) positional data are compared throughout 80% of the vowel interval. The time variable is vowel normalised time (0-1), in order to see if the predictive nature of TDz on *F*1, and TDx on *F*2, changes from one tongue configuration to another for each diphthong.

Formant trajectories were downsampled to the temporal resolution of EMA (100 Hz) using linear interpolation, in order to observe their relation as factors in mixed effects models, as well as how this relation changes over time.

4.5.3. Preliminary investigation: Results

Figures 11 (a – d) depict Smoothing Spline ANOVA (SSANOVA) models for TDz, *F*1, TDx, and *F*2 respectively over time (calculated from normalised values over normalised time). Linear mixed effects models were used to measure the predictive power of TDz on *F*1 and of TDx on *F*2, as reported in Table 2, below.

For TDz and *F*1: There was a significant effect for all diphthongs (*side*, *soid*, *sayed*, *sewed*, *sowd*), which is negative in all cases, thus in the expected direction. There was also

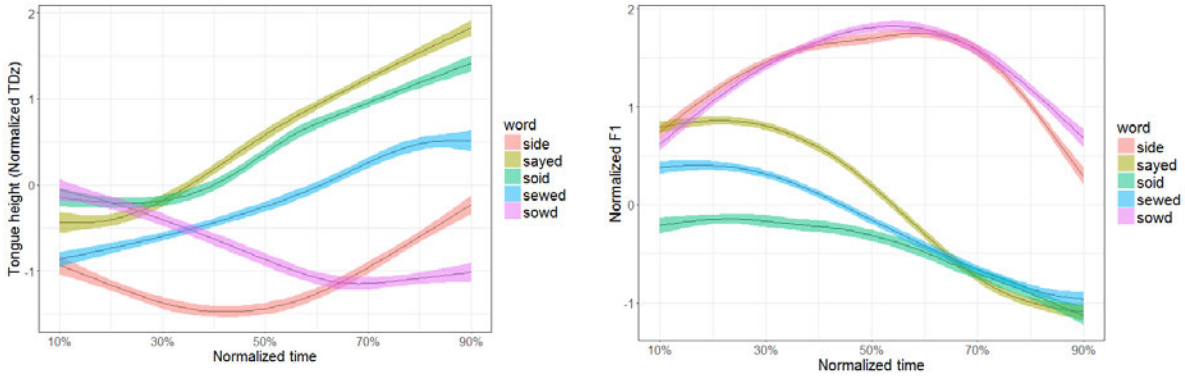
a significant interaction with time (normalised) for all diphthongs. The estimates are negative for *side*, *soid* and *sayed*, the rising diphthongs, and positive for *sewed* and *sowd*.

For TDx and *F*2: There was a significant effect for all diphthongs, except *sewed*. Interaction with time was also significant (again, *side*, *soid* and *sayed* were negative; and *sowd* was positive).

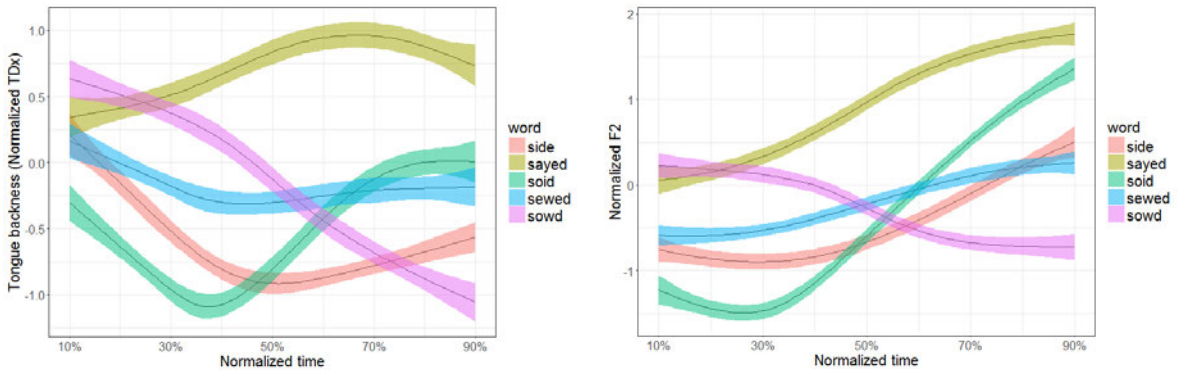
Table 2. Linear Mixed Effects models for F1 ~ TDz and F2 ~ TDx

***F*1 ~ TDz**	**Main effect**		**Interaction with time**	
Word	**Estimate**	***P*-value**	**Estimate**	***P*-value**
side (PRICE)	-0.68	$<2e^{-16}$	-0.10	$4.17e^{-05}$
soid (CHOICE)	-0.48	$<2e^{-16}$	-0.40	$<2e^{-16}$
sayed (FACE)	-0.90	$<2e^{-16}$	-0.66	$<2e^{-16}$
sewed (GOAT)	-0.71	$<2e^{-16}$	0.24	$3.3e^{-08}$
sowd (MOUTH)	-0.30	$6.46e^{-11}$	0.13	$<2e^{-16}$

***F*2 ~ TDx**	**Main effect**		**Interaction with time**	
Word	**Estimate**	***P*-value**	**Estimate**	***P*-value**
side (PRICE)	-0.07	<0.017	-0.32	$<2e^{-16}$
soid (CHOICE)	0.74	$<2e^{-16}$	-0.77	$<2e^{-16}$
sayed (FACE)	0.54	$<2e^{-16}$	0.69	$<2e^{-16}$
sewed (GOAT)	0.02	0.504		
sowd (MOUTH)	0.56	$<2e^{-16}$	0.20	$<2.19e^{-10}$



11a 11c



11b 11d

Figure 11. (a) and (b) display SSANOVAs for tongue height and tongue backness respectively. (c) and (d) display SSANOVAs for F1 and F2 respectively. Values are calculated from normalised articulatory and acoustic measurements over normalised time.

4.5.4. Preliminary investigation: Discussion

As with static measurements in monophthongs, TDz predicts *F*1 across the vowel interval in diphthongs. The strength of the linear relation between TDz and *F*1 varies across time (over the vowel interval). For some vowels (*sewed* and *sowd*) the relationship

becomes stronger over time and in others it becomes weaker (*side*, *sayed*, *soid*). TDx was a significant predictor of *F*2 across time for all vowels, except *sewed*, also in the direction expected. Thus for those vowels, (*side*, *sayed*, *soid*, *sowd*) as TDx advances, *F*2 measures increase, but the strength of the relationship also changes over time. Before speculating too much on the implications of these results finer grained analysis is required to determine where (in articulatory space) TDx becomes less predictive of *F*2. This was only an initial foray into this type of analysis, but shows promise.

In the main part of this thesis, most of the discussion was in relation to articulatory space (and tongue curvature), so it would be better if articulatory space were a factor in the analysis rather time. It is likely that the interaction with time in the models presented in this preliminary analysis were due to differences in articulatory space rather than time *per se*. This is not the only view that can be taken regarding future directions. For example, Gorman and Kirkham (2020) argue for a better understanding of time. Future work should look at developing ways of better quantifying time-varying acoustic-articulatory relations that can better take into account the complex gestural configuration that accompanies laterals and how this interacts with vowel gestures in various contexts (i.e. coarticulation).

Part of the question being investigated in this preliminary diphthong work is: Where in the vocal tract does the relationship between *F*2 and TDx change? The tipping point, or change in *F*2 by back-cavity-size-relation in a two-tube model occurs when the back cavity becomes longer than the front cavity (Stevens, 1989). Thus if all data samples (pairings of TDx and *F*2 for back vowels) were coded “back” (meaning the back cavity is smaller than

the front cavity) or "front" (meaning the back cavity is larger than the front cavity), then there should be an interaction between TDx and vowel backness (front versus back) when predicting *F*2. This prediction relies on the assumptions that (1) the two-tube model is applicable and (2) TD position is an appropriate approximation of how the front and back cavities are divided. This may not be valid for some diphthongs, such as onglides because the two-tube model is not valid throughout the whole vowel (due to the vowel moving towards a Helmholtz resonator).

Another way to test the hypotheses might be to compare individual speakers where known differences exist (e.g. F07 and M05, c.f. STRUT and FOOT in Figure 8). This, however, is the extent of this preliminary work and discussion. There is no answer as yet, there is more work still to do in testing the hypotheses more directly (with articulatory space rather than time). Diphthongs allow us to look at the question from another interesting angle.

4.6. Conclusions

At the beginning of this thesis, the question was posed: "Does an approach relying wholly on established theory about how acoustics map to articulation, i.e., examining only acoustic data, fully capture variation in performance? Or, is it necessary to collect articulatory data to obtain a full characterisation of a language or of the difference between dialects?" What has been shown here is that how one describes the vowel space depends on what type of data is examined, and this in turn influences conclusions made

about dialect differences. For two dialects, NAmE and AusE, we reported acoustic vowel spaces based on formants (*F*1 and *F*2) and articulatory vowel spaces based on the position of the TD. Several differences– such as the merger of LOT and THOUGHT in NAmE but not AusE – are reflected clearly in both types of data. Others are not. For example, the GOOSE vowel is central in both dialects if viewed articulatorily, but on the basis of *F*2, it is back in NAmE and central in AusE. In general, significant differences in TD position corresponded to significant differences in formant values, but the reverse was not always found. Differences in formants not reflected in TD position underscore the role that other aspects of articulatory control have on formant values. In particular, we demonstrated several cases in which lip rounding and tongue curvature (as a proxy for pharyngeal cavity size) plausibly perturb correspondence between TD position and formants.

With regard to the relationship between acoustics and articulation often assumed in dialect descriptions– namely, that *F*1 is inversely related to vowel height and *F*2 is inversely related to backness – we confirmed both this general trend as well as some predicted exceptions and individual differences. There were significant linear correlations across all of the data for *F*1 and TD height and weaker correlations for *F*2 and TD backness, but we also found that the strength of the linear relation was stronger for some speakers than for others and that it breaks down in some regions of the vowel space, e.g., low back vowels. We argue that both the general trend and the exceptions follow from the theoretical bases of the common heuristic. Thus, while formants are shaped by too many factors to be predicted by TD position alone and TD position cannot be accurately inferred from formants, the partial correspondence is highly encouraging. Not only does articulatory data enhance the description of dialect variation, but variation across dialects

offers an insightful probe into the relation between speech acoustics and articulation. Conducting experiments where direct observation of the articulators during a given speech sound can have implications not only for dialect research and speech science but also for other related fields such as speech pathology and language learning. I hope that this research has contributed in some way to the area in general.

References

Bernard, J.-B. (1970). A cine-X-ray study of some sounds of Australian English. *Phonetica, 21*(3), 138-150.

Berry, J. (2011). Accuracy of the NDI wave speech research system. *Journal of Speech, Language, and Hearing Research, 54*(5), 1295-1301.

Campbell-Kibler, K., Walker, A., Elward, S., & Carmichael, K. (2014). Apparent time and network effects on long-term cross-dialect accommodation among college students. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 4.

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

Chiba, T., & Kajiyama, M. (1941). The vowel: Its nature and structure. Tokyo: Tokyo-Kaiseikan Publishing Co.

Cox, F. (1999). Vowel change in Australian English. *Phonetica, 56*(1-2), 1-27.

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics, 26*(2), 147-179.

Cox, F. (2019). Phonetics and phonology of Australian English. In *Australian English Reimagined* (pp. 15-33). New York: Routledge.

Dawson, K. M., Tiede, M. K., & Whalen, D. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical linguistics & phonetics, 30*(3-5), 328-344.

Dromey, C., Jang, Sl, & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication, 55*(2), 315-328.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, *108*(1), 343-356.

Erickson, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, *59*(2-3), 134-149.

Fant, G. (1960). Acoustic Theory of Speech Production. The Hague: Mouton.

Foulkes, P., Scobbie, J.M., & Watt, D. (2010). Sociophonetics. In W.J. Hardcastle, J.L. Laver, & F.E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd Ed., pp. 703-754), Oxford: Wiley-Blackwell.

Gafos, A. I., Hoole, P., Roon, K., Zeroual, C., Fougeron, C., Kühnert, B., D'Imperio, M., & Vallée, N. (2010). Variation in overlap and phonological grammar in Moroccan Arabic clusters. *Laboratory Phonology X, Mouton de Gruyter, Berlin/New York*, 657-698.

Georgeton, L., Kocjančič, T., & Fougeron, C. (2014). Domain initial strengthening and height contrast in French: Acoustic and ultrasound data. In *10th International Seminar on Speech Production* (pp. 142-145).

Gorman, E., & Kirkham, S. (2020). Dynamic acoustic-articulatory relations in back vowel fronting: Examining the effects of coda consonants in two dialects of British English. *The Journal of the Acoustical Society of America*, *148*(2), 724-733.

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.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/ fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*(5), 2825-2835.

Harrington, J., Kleber, F., & Reubold, U. (2011). The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association*, *41*(02), 137-156.

Hermann, L. (1890). Phonophotographische untersuchungen. *Archiv für die gesamte Physiologie des Menschen und der Tiere* 47, 44-53.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America, 97*(5), 3099-3111.

Holmes, J., & Hazen, K. (2013). *Research Methods in Sociolinguistics: A Practical Guide.* (1st ed., GMLZ - Guides to Research Methods in Language and Linguistics). Somerset: Wiley.

Iskarous, K. (2001). *Dynamic acoustic-articulatory relations*. Ph.D. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL.

Iskarous, K. (2010). Vowel constrictions are recoverable from formants. *Journal of Phonetics, 38*(3), 375-387.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical society of America, 94*(2), 701-714.

Kirkham, S., & Wormald, J. 2015. Acoustic and articulatory variation in British Asian English liquids. *Proceedings of the XVIII International Congress of Phonetic Sciences* 1-5.

Kochetov, A. (2020). Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass*, *14*(4), 1-1.

Koenig, W., H.K. Dunn, H.K., & Lacy, L.Y. (1946). The sound spectrograph. *Journal of the Acoustical Society of America* 18(1), 19-49.

Lammert, A., Proctor, M., & Narayanan, S. (2013). Interspeaker variability in hard palate morphology and vowel production. *Journal of Speech, Language, and Hearing Research*, *56*(6), S1924-S1933.

Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the Acoustical Society of America*, *146*(6), 4363-4381.

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America*, *139*(1), 426-440.

Leung, Y., Oates, J., Papp, V., & Chan, S. P. (2020). Formant frequencies of adult speakers of Australian English and effects of sex, age, geographical location, and vowel quality. *Journal of Voice*. (in press, corrected proof)

Lin, S., Palethorpe, S., & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. In F. Cox, K. Demuth, S. Lin, K. Miles, S. Palethorpe, J. Shaw & I. Yuen (Eds.), *Proceedings of the 14th Speech Science and Technology Conference* (pp. 105-108). Sydney: Australasian Speech Science and Technology Association.

Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4B), 1166-1179.

Lobanov, B. (1971). Classification of Russian vowels spoken by different listeners. *The Journal of the Acoustical Society of America. 49*:606-608.

McGowan, R.S., & Berger, M.A. (2009). Acoustic-articulatory mapping in vowels by locally weighted regression. *Journal of the Acoustical Society of America, 126*(4), 2011-2032.

Noiray, A., Iskarous, K., & Whalen, D. (2014). Variability in English vowels is comparable in articulation and acoustics. *Laboratory phonology, 5*(2), 271-288.

Nycz, J. (2013). New contrast acquisition: methodological issues and theoretical implications. *English Language and Linguistics*, *17*(02), 325-357.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel/u: A pilot ''motor equivalence'' study. *The Journal of the Acoustical Society of America*, *93*(5), 2948-2961.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical society of America, 24*(2), 175-184.

Potter, R.K., Kopp, G.A., & Green, H.C. (1947). *Visible Speech*. New York: Van Nostrand.

Ratko, L., Proctor, M., Cox, F., & Veld, S. (2016). Preliminary investigations into the Australian English articulatory vowel space. *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology, December 2016*, Parramatta, Australia. 117-120.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, *1*(4), 333-382.

Scobbie, J. M., Lawson, E., & Stuart-Smith, J. (2012). Back to front: a socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Rivista di Linguistica/Italian Journal of Linguistics, Special Issue: "Articulatory techniques for sociophonetic research"*, *24*(1), 103-148.

Shaw, J.A., Chen, W.R., Proctor, M.I., & Derrick, D. (2016). Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech, Language, and Hearing Research, 59*(6), (pp. 1566-1574*).*

Shaw, J.A., Tyler, M.D., Kasisopa, B., Ma, Y., Proctor, M.I., Han, C., Derrick, D., & Burnham, D.K. (2013). Vowel identity conditions the time course of tone recognition. In *INTERSPEECH,* (pp. 3142-3146).

Sóskuthy, M., Foulkes, P., Haddican, B., Hay, J., & Hughes, V. (2015). Word-level distributions and structural factors codetermine GOOSE fronting. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*, UK: The University of Glasgow. (pp. 1001-1006).

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics, 17*, 3-45.

Stone, M., & Vatikiotis-Bateson, E. (1995). Trade-offs in tongue, jaw, and palate contributions to speech production. *Journal of Phonetics*, *23*(1), 81-100.

Strycharczuk, P., & Scobbie, J. M. (2017). Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America*, *142*(1), 322-331.

Tabain, M. (2008). Production of Australian English language-specific variability. *Australian Journal of Linguistics, 28*(2), 195-224.

Tilsen, S., & Goldstein, L. (2012). Articulatory gestures are individually selected in production. *Journal of Phonetics*, *40*(6), 764-779.

Turton, D. (2017). Categorical or gradient?: An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology 8*(1), 1-31.

Urbassek, H. M. (2014). Acoustic tube model of the human vocal tract: formants and vowels. *European Journal of Physics*, *35*(4), 045017.

Vilain, C., Berthommier, F., & Boë, L. J. (2015, September). A brief history of articulatory-acoustic vowel representation. In *HSCR 2015-1st International Workshop on the History of Speech Communication Research*.

Watson, C. I., Harrington, J., & Palethorpe, S. (1998). A kinematic analysis of New Zealand and Australian English vowel spaces. Paper presented at the *5th International Conference on Spoken Language Processing.* (Vol 6), (pp. 2363-2366).

Watt, D., & Tillotson, J. (2001). A spectrographic analysis of vowel fronting in Bradford English. *English World-Wide, 22*(2), 269-303.

Wells, J. C. (1982). *Accents of English.* (Vol. 1), (pp. 117-183). Cambridge: Cambridge University Press.

Whalen, D. H., Chen, W. R., Tiede, M. K., & Nam, H. (2018). Variability of articulator positions and formants across nine English vowels. *Journal of phonetics*, *68*, 1-14.

Wieling, M., & Tiede, M. (2017). Quantitative identification of dialect-specific articulatory settings. *The Journal of the Acoustical Society of America*, *142*(1), 389-394.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics*, *59*, 122-143.

Wood, S. (1982). *Working Papers 23: X-Ray and Model Studies of Vowel Articulation*. Lund: Lund University Department of Linguistics.

Appendices

APPENDIX 1.

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom
***F*1:** *F*1 ~ vowel * dialect + (1 | speaker) + (1 | order)

	Sum Sq	**Mean Sq**	**Num *DF***	**Den *DF***	***F*-value**	***P*-value**
Dialect	0.00	0.002	1	633	0.03	0.8687
Lexical Set	556.84	61.871	9	633	808.24	<0.001
Accent:Lexical Set	34.49	3.833	9	633	50.07	<0.001

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom
Tongue Dorsum Height

	Sum Sq	**Mean Sq**	**Num *DF***	**Den *DF***	***F*-value**	***P*-value**
Dialect	0.00	0.003	1	6.99	0.03	0.8769
Lexical Set	403.34	44.816	9	624.96	395.98	<0.001
Accent:Lexical Set	31.11	3.457	9	624.96	30.54	<0.001

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom
***F*2**

	Sum Sq	**Mean Sq**	**Num *DF***	**Den *DF***	***F*-value**	***P*-value**
Dialect	0.002	0.0025	1	633	0.005	0.9434
Lexical Set	153.172	17.0191	9	633	34.409	<0.001
Accent:Lexical Set	167.814	18.6460	9	633	37.699	<0.001

Analysis of Variance Table of type III with Satterthwaite approximation for degrees of freedom
Tongue Dorsum Backness

	Sum Sq	**Mean Sq**	**Num *DF***	**Den *DF***	***F*-value**	***P*-value**
Dialect	0.000	0.0001	1	6.96	0.005	0.9467
Lexical Set	36.756	4.0840	9	624.67	172.951	<0.001
Accent:Lexical Set	9.026	1.0029	9	624.67	42.473	<0.001