UNIVERSITY OF CANTERBURY

DOCTORAL THESIS

Authentic Self, Incongruent Acoustics: A Corpus-Based Sociophonetic Analysis of Nonbinary Speech

Author:
Jacq JONES

Supervisors:
Dr. Jennifer HAY
Dr. James BRAND
Dr. Erin HARRINGTON
Dr. Viktoria PAPP

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Linguistics

at the

University of Canterbury

November 4, 2022

"It'd be a poor kind of world where there was just one explanation for things."

Diane Duane, *To Visit the Queen*

Abstract

Jacq JONES

Authentic Self, Incongruent Acoustics: A Corpus-Based Sociophonetic Analysis of Nonbinary Speech

This thesis examines the ways six nonbinary speakers in Christchurch, New Zealand present their gender identity via speech. It examines their productions in reference to both established trends in the literature, as well as speech collected from ten binary speakers (5M, 5F) at the same time. It seeks to examine whether, in addition to encoding binary gender, speech also encodes *nonbinary* gender.

Three hypotheses are proposed and tested across multiple linguistic variables. The first hypothesis regards acoustic *incongruence*, and posits that nonbinary speakers may assert their nonbinary identities via speech that utilises particular combinations of variables which create either *ambiguity* or *dissonance* in regards to established binary-gender norms. Ambiguous gender incongruence arises from the use of speech that is neither reliably perceived as female, nor reliably perceived as male. Dissonant gender incongruence arises from the use of speech that is reliably perceived as both male and female. The second hypothesis predicts that nonbinary speakers will show greater variation in speech based on immediate contextual factors, compared to binary speakers. This difference is hypothesised to be due to to nonbinary speakers paying greater attention to production, and the greater degree of variation in their own speech over time compared to binary speakers. Hypothesis 3 predicts that nonbinary speakers are not a uniform population, and that their use of incongruence will be influenced extensively by their individual condition, including their professed speech goals, history, and gender identity.

The hypotheses are tested quantitatively in regards to five linguistic variables: Pitch, pitch range, monophthong production, Vowel Space Area (VSA), and intervocalic /t/ frication rates. The interaction between multiple variables together is also considered. In-depth examinations of the variation utilised by a single speaker in the form of "Spotlights" address the hypotheses from a qualitative perspective.

Overall, the thesis finds some evidence for Hypothesis 1. In every linguistic variable examined, nonbinary speakers show some distinction from binary speakers that is not explained fully via speaker Assigned Sex at Birth (ASAB). Some binary speakers also seem to produce incongruence, particularly binary women and particularly within single variables. The small scale of the study presents a limitation in addressing Hypothesis 2, but avenues for future work are identified. The qualitative evidence provides strong support for Hypothesis 3, in the examination of individual nonbinary speakers and the way their measured productions support their professed speech goals and identities. Overall, this dissertation presents one of the first comparative analyses of nonbinary speech, and presents a number of novel approaches to examining phonetic data from a statistical perspective that still accommodates an analysis of individual agency and goals in identity building.

Acknowledgements

There are so many people whose support was invaluable throughout this process. Without them, this dissertation would be terrible, or would not exist at all. Without participants this research wouldn't exist. I am grateful for your time and your words, and hope I have done right by you. I wish to acknowledge the following for the academic, financial, and pastoral support: The UC Doctoral Grant, LSAP funding, the Three-Minute Thesis competition, the PGSA, and the rainbow support networks provided by UC and the TEU.

My supervisors provided advice, kindness, patience (SO much patience), and endless support. Erin, for insight and support in cultural studies, pastoral support, and feminist yarns. Vica, for coding expertise and practical advice. James, for statistical (and, let's be real, emotional) support and extensive manuscript feedback. Jen, my primary supervisor, for providing all of the above and so much more, who taught me more about sociolinguistics, statistics, academia, and real, genuine, compassionate curiosity than I can begin to put into words. Your guidance has helped me grow as a scholar and a human. Thank you all so much for everything.

Lynn and Susan provided invaluable professional and academic support throughout my PhD, and were also good friends and cheerleaders. Jeanette, Kevin, Jonathan, Donald, and all the other Socio Slot regulars, who provided invaluable feedback, support, and snacks over the years. The other Corridor Crew: Heidi, Dineke, Megan, Yoon Mi, Clay, Simon, Josh, Forrest, Haipeng, Muhammad, Jiao Dan, Andy, Matthias, Wakayo, Darcy, Keyi, Xuan, Ryan, Daiki, Moonsun, Muneir, Mineko, Usma, Marie, Stephanie, Sidney, Vicky, Matt, Hayden, Haipeng, and Peiman. Visiting scholars Sarah Hawkins and Jane Stuart-Smith provided new perspectives and rabbit holes. Robert deserves props for technical support, general nerdery, and never making me feel like I wasn't nerdy enough. Last but not least, Emma always had the BEST chats, check-ins, and more.

I'm grateful to my family, who have always spread themselves around the world but never lost touch, which has been helpful in navigating the new normal. I can't wait until covid allows us to travel. I miss you all so much. My partner Andrew, who has put up with so much to help me through, and supported me at every step, even the steps backwards. Finally, Io, who didn't help at all, not even a tiny little bit.

The Cute Officemates, the Thirsty Sword Lesbians, the Despondent Cowboys, and the Thursday Night Crew. I couldn't ask for a better group of friends to spend time with. Thanks to Dan for being the biggest help ever, not only in terms of feedback and stats-wrangling, but also being an endless well of compassion, humour, and insight. Cyn for being an early and constant lover of language and source of support. Jennifer for being an excellent friend and person to bounce ideas off of. Thanks pals, for the heists, the stories, the spooks, and the groans. I also know that I've forgotten at least one incredibly valuable person, and will remember them the moment after I submit this document. I apologise in advance for leaving you out.

I could not have completed this dissertation without all of the animals that allowed me to pet them more than once: Clover, Lopan, Rory, Ziggy, Marley, Biscuit, Nubbin, Parsnips, Poppy, and Scorpion. I hope I get to pet you all more in the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction to the Introduction

This dissertation is an examination of the sociophonetics of nonbinary gender. It takes a holistic approach in examining the speech of nonbinary people in Ōtautahi Christchurch, Aotearoa New Zealand, in relation to pervasive social stereotypes about sex/gender, the speech of their binary-gendered peers, and their personal attitudes and relationships to gender. It achieves this examination via the creation of a corpus of Recorded Audiovisual Interviews with Nonbinary and Binary Orators , the RAINBO corpus, which has been designed with the intention of sustainability and future growth.

In this introductory section, I provide some brief (but necessary) context to situate the research. First, I introduce the concept of genders beyond the established binary of man/woman, and nonbinary people as a research population in Section 1.1.1. Having established this groundwork, I then present my broad research questions in Section 1.2. In Section 1.3, I provide an overview of the particular terminology I have chosen to use, and my approach to disclosure throughout this work. I then outline the structure of the rest of the thesis in Section 1.4.

1.1.1 Defining Nonbinary

When a child is born, the delivering doctor runs through a checklist: Ten fingers, ten toes, pumping lungs, beating heart, eyes and mouth and nose and private parts. In most cases in popular conception, the delivering parent then hears a proud exclamation, “It’s a boy!” or “It’s a girl!” This is an idealisation of a process that involves the assignation of gender, and with that assignation, sometimes before being given a name (or, increasingly, before birth), a roadmap is laid for the child’s life, from their upbringing during their formative years, their opportunities and access to education, wealth, and power, and to their expected roles within the context of reproductive labour (Secombe, 1974; Engels, 2021). From birth, these social expectations of sex/gender, of what it means to “talk like a boy” or “play like a girl”, are reinforced (Witt, 1997, p. 255). Even very young children are exposed to the expectations inherent in these roles, if only for the purpose of recognising transgression e.g. *"she’s* not like the other girls, she likes trucks!".

However, and thankfully, like most things we do as infants, it is possible to grow out of, or beyond, the sex/gender designation received at birth. The term *transgender* (sometimes shortened to *trans* or used adjectivally as in *trans women* or *transmasculine*) refers to this lack of resonance with one’s Assigned Sex at Birth (ASAB). For some trans people, a rejection of their ASAB involves a confirmation of the alternate binary gender. For others, this discomfort with ASAB is less about the rejection of one binary option in favour of the other, but potentially a resonance with both options, or neither,

or something in between, or something else completely, or sometimes even a rejection of the entire system upon which binary gender is scaffolded. This constellation of *nonbinary* identities is the population I am concerned with in this thesis: People whose gender does not align with either of the culturally-entrenched binary options of man/male or woman/female.

As in the idealised "delivery room" scenario described above, in popular culture, mass media, and the academic literature, it is taken as given that gender is a binary, or at best a binary spectrum. Many people who collect population-level information want simply to know the best and most efficient way to analyze (and usually market towards) the majority. In the digital age where ads are served to users based on abstracted search data and algorithms, it can be as simple as wanting to know whether a potential customer is more likely to buy a razor in a pink package or a blue one (Hogg and Garrow, 2003; Kim, 2017). Reinforcing a stereotype that leads to "men use blue razors" and "women use pink razors" makes targeting those ads more efficient. The pragmatism behind an enforced binary is not limited to those with commercial interests. The most recent Aotearoa New Zealand census, in 2018, contained a mandatory question about sex/gender that gave only the binary options of “male” and “female”. Since the previous census in 2014, there had been a continuous discussion across all media for changes to include more diverse sex/gender identities. In a best practices document released in 2016[1], Statistics New Zealand states “In order to measure the diversity and complexity of gender identity (physical, genetic, emotional, etc) and sexual orientation concepts, new and separate definitions, questions, classifications, and statistical standards would be required.”(Statistics New Zealand, 2016, p. 1).

Despite a self-congratulatory announcement in 2015 that "world-first" gender diverse options would be added to the 2018 census (Statistics New Zealand, 2015), in August 2017 the decision was made to quietly scrap these questions. According to online news magazine *Stuff*'s coverage at the time (McDonald, 2017), "statisticians could not make [gender diverse options] work." Changing the questions regarding sex and gender identity would lead to inaccurate, inconsistent, and unusable data, according to the statistician interviewed for the article, as test surveys indicated Statistics New Zealand would be “unable to guarantee the quality of the information” due to what they judged to be flippant or inaccurate answers.[2] However practical these considerations, they serve to reinforce the binary ideology, and this assumption can be seen across multiple domains, including within the linguistics literature (which I discuss more specifically in Chapter 2.2.1). Nonbinary people, in this conception, are a statistical minority that introduce more complications than their accurate accounting is “worth”. This omission, in turn, reinforces the erasure and marginalisation of nonbinary identities, both in practical terms (one goal of national censuses is to target social programs and spending to populations that need it), and in popular thought, as the body responsible for counting members of a population is also its gatekeeper. This problem is not unique to gender: Similar issues have elicited comment concerning the same census' ethnicity questions, which have separate indications for “European” and “New Zealand European” but not “New Zealand <other ethnicity>”. This choice is contextualised historically by Aotearoa New Zealand's unofficial “white” immigration policies (Ward and Masgoret, 2008, p. 228) that elevates immigrants of

[1] Media releases and data available online prior to the 2018 census have been removed from the Statistics New Zealand website and online archives since at least December, 2020, but are still available via direct request from Statistics New Zealand (Mair, 2021).

[2] It is not clear from the article whether these indicate "joke" answers from cis binary participants, genuine answers from nonbinary participants Statistics NZ found objectionable, both, or some other issue entirely.

European descent, but alienates and erases those of other ethnicities with Aotearoa New Zealand roots, as in UMR's case study of a fourth-generation Chinese New Zealander (UMR Research Ltd., 2016, pp. 154–158).

Even in the context of these institutional barriers and erasure, worldwide the numbers of people identifying themselves as nonbinary is growing, particularly in the Global North. A survey conducted by the HRC (Human Rights Campaign, 2013) of more than 10,000 Lesbian, Gay, Bi, Trans, and Others (LGBT+) youth aged 13-17 in the USA, 6% identified as some form of nonbinary: By 2018, a follow-up study found this number had increased to 26% (Human Rights Campaign, 2018, p. 24). Titman (2014), using census data from the UK, estimates about 0.4% of respondents to be nonbinary, though Titman's methodology involved imputing the numbers based on survey responses where the binary gender options were omitted or defaced. A survey conducted by Ipsos in 2021 across 27 countries found 4% of "Gen Z" (respondents born in 1997 or later) identifies as something other than male or female, compared to 1% of other adults (Ipsos, 2021).

These changes in the conception of gender are not exclusive to LGBT+ populations, either. A survey conducted in Israel among normative cis populations (those who identified themselves as "male & man" or "female & woman"), found that more than 30% gave nonbinary-coded responses (such as feeling like a gender other than one's ASAB or experiencing dysphoria) (Joel et al., 2014). This latter result, taken with the increase over time found by the HRC, may indicate a growing dissatisfaction with the sex/gender binary as a whole, even while awareness that there are other options remains low. Locally, in Aotearoa New Zealand, The Counting Ourselves Survey specifically targeting trans and nonbinary respondents (Veale et al., 2019) found that 40% of respondents positively identified as Non-Binary.[3] In all three studies (Joel et al. (2014), HRC, and Counting Ourselves), the number of nonbinary responses increases dramatically when multiple options for identifying one's sex/gender are given, further highlighting the gatekeeping effect inherent in traditional survey designs, like the census.

1.2 Research Questions

In the previous section, I have described nonbinary gender and established that nonbinary people are a relevant and determinable population. I have also presented evidence from survey data that, as a population, their numbers appear to be a growing (at least in the context of the English-speaking Global North), despite evidence that there is a continued social and institutional pressure to erase and subsume them in favour of a continued adherence to a sex/gender binary. With this data, and the context that this is a thesis on sociophonetics, the overarching intention of this thesis is to answer to the following research questions:

1. Given their status as an emergent, distributed, and minority population, does nonbinary speakers' use of gendered speech features differ from that of binary speakers due to their nonbinary gender?

[3] This survey also used "Non-Binary" as a catchall term for respondents who did not ID exclusively with femme-aligned (woman, trans woman, tangata ira wahine, whakawahine) or masc-aligned (man, trans man, tangata ira tāne) terms, in which it found that the Non-Binary category increased to 45% of respondents (with trans men accounting for 29% and trans women the remaining 26%.

2. Given the pervasive historic and enforced gender binary, to what extent do existing methods of sociophonetic analysis support the inclusion of nonbinary speakers?

These research questions are necessarily broad, as in any examination of a relatively new (or newly recognised) population. I do not intend to answer them fully or completely, but to treat them as a guide and a reminder of the context of nonbinary people as juxtaposed to established gender stereotypes. The first question seeks to establish a deceptively straightforward yes/no response – Are they different? Are these differences such that we can expect them to manifest in speech? As I will explore in the following chapter, many of the things we know about variation in speech make assumptions about physiology and community practice that are somewhat inappropriate or inapplicable when applied to nonbinary people. Simply, the initial goal of the first research question is to identify whether significant variation is found between the speech of nonbinary speakers and the speech of binary men and women, in a way clearly attributable to their gender.

The second research question takes a more reflexive point of view, focussing on the ways the gender binary has shaped established methodologies and narratives within sociolinguistics. If the first research question addresses what traditional methods of acoustic analysis can teach us about emergent identities, the second addresses what an examination of those identities can teach us about the traditional conceptions and analyses.

The hypotheses I present in the following section are likewise not attempts to answer the research questions fully and completely, but to narrow their focus further still, to questions that are actionable and testable within the scope of this dissertation. While the hypotheses and the research itself were conceived with the intuition that observable differences between nonbinary and binary speakers exist, the strategies employed in addressing the research questions do not change if no significant differences are found. If nonbinary speakers are indistinguishable from binary speakers, why not? A number of possibilities emerge: The scarcity of peers due to the distributed, emergent status of the population may make variation difficult to implement or reinforce. It might be for lack of trying, if the assumption is that nonbinary identity is external or unaffected by the voice. Or, it might be a deficiency in the methodology to identify the kinds of changes involved in binary vs nonbinary production. The theoretical frameworks regarding identity management, stance-taking, articulatory settings, and the role of past experience make predictions about observed similarity equally as capably as observed difference.

1.3 Notes about Terminology & Disclosure

1.3.1 Terminology

In addition to the institutional processes described above that make a full accounting of nonbinary populations difficult, there are a number of other confounds inherent in the way we describe sex/gender that make it difficult to talk about nonbinary people and non-binary experiences. The concepts of “sex” and “gender” are not defined equally or concretely across contexts and authors. When a distinction is emphasised, *sex* is generally conceived of as appealing to a biological absolute, a collection of features that clearly separates the human population into two groups, "men" and "women." These features are somewhat variable but tend most often to include distribution of sex chromosomes (women are XX, men are XY), hormone

dominance (men are testosterone-dominant, women are estrogen-dominant), gamete production (women produce large ovules and men produce small spermatozoa), and physical conformation (genitals, body size, location of hair, etc). These features are variably weighted, both in regards to how "important" each is in determining sex (for example, as of 2021 the International Olympic Committee is concerned only with hormone balance (International Olympic Committee, 2021)), and by the degree to which variation within and across each feature is seen to determine sex proportionately.

In practice, "sex" refers most commonly to a person's ASAB, which is determined by the delivering doctor based solely on the appearance of the child's genitalia (Hester, 2004, p. 25). Of course, about 1 in 2000 people will have a biology at birth that does not conform to binary sex standards.[4] For context, in terms of rates of appearance and the subsequent societal concern and treatment, this would be similar to a culture choosing to medically intervene in (or ignore the existence of) people with red hair (Mitchell, 2018).[5]

The definition of *gender*, as opposed to *sex*, is usually based on an appeal not to biology, but to broader social ideologies about confirmation or deviance: That is, what do men and women *do*, look like, etc, that differentiates one from the other. However, the definitions are equally as fuzzy as the biological ones presented earlier. They rely on a number of interpretations and assumptions regarding which society forms the context, what behaviors are considered deviant (or, at least, deviant enough to police), what behaviors are considered gender-conforming (or conforming enough to celebrate), and whether to make a distinction between internal gender identity, external gender expression, and the relationship of both these things to one's physical appearance or ASAB.

Sex may be framed as a strictly objective definition of biological feature bundles, but in practice still relies on a social interpretation of a person to infer those biological features. In the same way, our impressions about gender as a social concept are informed by our assumptions about the underlying biology. I present two definitions of gender here, from different authors, that illustrates the complex, contextual, relative nature of gender as a category assigned to humans.

> "For the sake of definitional clarity, let us take it that by gender culture is meant a society's understanding of what is possible, proper, and perverse in gender-linked behavior, and more specifically, that set of values, mores, and assumptions which establishes which behaviors are to be seen as gender-linked, with which gender or genders they are to be seen as linked, what is the society's understanding of gender in the first place, and, consequently, how many genders there are." (Ramet, 1996)

[4] This estimate, 0.05%, is at the lower end of the scale and refers to a relatively narrow spectrum of intersex conditions with mostly obvious physical manifestations at birth. According to Hester (2004, p. 24), the number of children born with any kind of intersex condition is closer to 1.7%. This number is even higher if you also include changes to the endocrine system *after* birth, such as PCOS (polycystic ovary syndrome, where too much insulin and testosterone is produced by the body) or hypopituitarism (where some hormones cease to be produced by the body at all).

[5] Another point of caution arises here, because intersex people do not always identify as trans, and their experiences of marginalisation, oppression, and their relationship to and feelings regarding medical intervention are quite different to that experienced by many trans people. It is true that the existence of both trans and intersex people challenge the hegemonic binary conception of sex/gender (Costello, 2016), but that does not make them the same. That the two groups are often incorrectly conflated is further testament to the difficulty inherent in the language we use, as we lack ways to plainly separate different sources of opposition to the hegemonic gender binary.

> Gender is not considered property but "a relationship which brings about redefinitions of subjectivities and subject positions over time, both as products and as producers of social context" (Calás and Smircich, 1996, p. 241). "Subjectivity" is a sense of who you are. A "subject position" is a sense of how you are positioned in relation to others. Both are affected by or constructed through gender, which is not something residing inside the human, but a relational concept, just like 'big' cannot be 'big' unless there is something other than 'big' that makes it so (Gherardi, 1995). (Ahl, 2002, p. 47)

Finally, while not a direct concern within this work, *sexuality* and *attraction* are likewise intrinsically linked to sex/gender, as our language for describing the kinds of person someone is attracted to inherently relates to the relationship between that person's body and one's own.

A theme I return to throughout this thesis is the notion that the label nonbinary is not merely a third gender option, meant to stand alongside "men" and "women", but is an umbrella term that includes very disparate identities. For example, a non-exhaustive list of identities included under the nonbinary umbrella: Those who identify as *bigender* (both male and female), *pangender* (male and female and something else), *agender* (neither male nor female nor anything else), and *genderfluid* (moving between different genders at different points in time).

Unsurprisingly, the definition of "trans" (and related terms like *transgender* and the more outdated *transsexual*) can vary depending on the context and goals of the speaker, as well. The definition I use here is the most expansive, and contrasts with another popular definition that is much narrower. Sometimes referred to as a *transmedicalist* viewpoint, the narrow definition requires that, for a person to be trans, they must experience both a particular kind of "clinically significant distress" known as gender dysphoria, and, by implication, a desire to access medical interventions to reduce that distress (American Psychiatric Association, 2013). Transmedicalist definitions tend to exclude some, or all, nonbinary identities, and proponents of the transmedicalist definition often characterise nonbinary identities as invalid, using insulting terms like *theyfab* (a play on Assigned Female at Birth (AFAB)) or *transtrender*.[6]

Within this thesis, I use "sex", "gender", and "sex/gender" interchangeably, to refer to a socially iterated sense of identity, with the understanding that sex is generally used to refer to the roles and assumptions that begin from the time one is assigned a sex at birth, as well as physiological characteristics of presentation, while gender is more often associated with individual agency, referring to social considerations and perceptions of one's identity, personality, etc. It is important to point out that popular terminology (and in fact understanding) takes "sex" as something more innate, unchangeable, and somehow more *real* than gender, and there can be a tendency to consider gender as something "overlaid" on sex, like clothing. Even while I explicitly collapse this definition formally here, I still lean upon this popular understanding, particularly within the participant interviews.

My definition does include a number of exclusions, though. I use nonbinary to refer to people whose gender is outside of the binary designation of "man" or "woman", excluding those whose identity is part of a culturally entrenched third

[6] It is interesting to note that this stereotype of a nonbinary person as young, confused, and seeking attention is the same one frequently levelled at bisexual women (Hertlein, Hartwell, and Munns, 2016).

gender[7] , and those whose identity is nonhuman[8]. I have included a glossary with a full list of definitions and abbreviations, which can be found in Appendix A.

I use *nonbinary* deliberately without a hyphen. While we (so far) lack the appropriate words to talk about nonbinary identities as a whole without reference to the gender binary, removing the hyphen is an attempt to somewhat present the word as a cohesive, self-contained unit, referring to an identity embodied by people as a coherent whole in itself, as opposed to a negation <*non-*> applied to a root <*-binary*>. Nonbinary is also more visually similar to other gendered terms like man, women, male, (after all, we do not write wo-man or fe-male). My personal impression is that *non-binary* encodes an explicit rejection of the binary, while *nonbinary* is an affirmative identity. Anecdotally, via an informal poll, this impression was shared by some of my nonbinary (and non-binary) friends, but not others: Some read the hyphen as more clinical or "more cis", while others did not, and self-identification showed rates that were about equal, with half using the hyphen and half not, for reasons varying from the introspective/stance discussed above, to the perceived formality of the hyphen, to no reason at all. In the rare cases I use *non-binary*, it either quotative (reflecting the exact use of a participant), or is intended to refer to a rejection of binary norms that does not necessarily imply nonbinary gender. For example, <*nonbinary speech*> is speech produced by a nonbinary speaker, while <*non-binary speech*> is speech that does not conform to established gendered norms, regardless of the gender of the speaker.

1.3.2 Disclosure

Something conspicuously absent in the following analysis, relative to other sociophonetic studies on gender (and more broader scholarship with trans people), is the explicit, repetitive disclosure and consequential centering of each participant's Assigned Sex at Birth. This shift in focus is a decision I made early on in the research process and one I have maintained throughout the course of writing this dissertation. It is often the case, in the course of this research (such as during conference presentations), that people will ask about a specific participant's ASAB.[9] It takes many forms, from well-meaning curiosity, to a kind of ASAB-guessing game, to (once) "But what are they *really*?". The questions are (almost) always given in a good faith, and in many cases are used as a shortcut to all of the associations we have about sex/gender in biology and society when we refer to "men" and "women" – For example, physiological factors like length of vocal folds, genital configuration, height, or hormone dominance, and social factors such as whether someone may have attended an all-girls' high school or been "socialised" a certain way with regards to sport. While I address this in more detail in the literature review in the following chapter, suffice it to say here that the answers implied by disclosing ASAB are not *actually* answered by ASAB alone.

I give the following justification for this deliberate masking:

[7] I excluded these unequivocally nonbinary populations because the limitations of my methodology did not allow me to create a point of comparison. Put simply, I am working under a paradigm in which nonbinary speakers are working within a presumed gender binary, which is not the case for those whose upbringing and experience contains knowledge and acknowledgment of other/third genders as valid. This exclusion was a mistake, and something I discuss more in Chapter 10.

[8] It is beyond the scope of this research to consider those whose identity is partially or wholly nonhuman, such as otherkin (Kirby, 2016). Current linguistic research does not recognise any nonhuman animal (or nonanimal) communication as language, and so it is not possible to measure the effects of this identity on speech production.

[9] Though I should note it has only ever been asked in reference to the nonbinary participants.

1. **Ethical** I do not believe this information is necessarily ethical to disclose, in regards to the paperwork submitted and subsequently approved by the ethics committee. As this study is on the ways identity (not ASAB) interacts with speech, it is appropriate to organise participants based on self-determined identity.

2. **Practical** Nearly all existing studies on sex/gender in the literature do *not* ask participants to disclose their assigned sex at birth, but rather rely on self-determination of sex/gender. This is especially true of cis participants. The information I collected from participants was their gender (via recruiting materials specifically seeking men, women, and nonbinary participants), the demographic questionnaire (which can be found in Appendix B) where participants were given a blank line to fill in their gender, and finally via the sociolinguistic interview where I asked "What is your gender?"

3. **Unknown** I did not directly ask any participants (non-binary or binary) to disclose to me their assigned sex at birth. This was deliberate for the reasons listed in 1. In some cases, participants disclosed their ASAB to me over the course of the interview, but in many cases they did not. Put simply, I cannot confidently report what I do not know.

Despite this lack of precise focus, it is also not my intention to hide information that is useful for analysis, or to present the data in a way that is intentionally misleading. In fact, the lack of disclosure is an attempt on my part to *minimise* confusion through repeated, intense focus on nonessential information. I am absolutely sure that, for every participant in this thesis, ASAB can be inferred directly from the information and quotes provided in reference to the participant bios in Chapter 5. In some cases, where I believe it is helpful for an analysis, I disclose ASAB obliquely (such as pointing out when two participants have the same or different assigned sexes at birth). So while I do not make this information clearly referable via a table, I do not disguise it.

In decentering ASAB my intent is to avoid its use as an axiom - the underlying "truth", starting point, or reality which all else is built on top of. My participants (and myself) exist as human beings, their (our) identities as women, nonbinary people, and men are not predicated on nor subsumed within their ASAB. If, in the course of reading this work, you find yourself engaged in the guessing game, scrutinising the data and quotes for some hint, some moment of epiphany at having "deduced" a participant's ASAB, I would like you to pause. I would like you to ask yourself where that curiosity is rooted, what purpose that surveillance serves, and what it might be feel like to be so audited when going about your daily life.

1.4 Structure of the Thesis

Chapters 2 through 5 establish the research questions, methodologies, and participants. Chapter 2 presents a contextual review of the literature, and outlines the gap in sociophonetics that the research questions and hypotheses seek to address. To address the research questions, in Chapter 3 I present my hypotheses regarding the use of gender incongruence to present nonbinary identity, and the assumptions that underlie the hypotheses. Chapter 4 outlines the research methodology involved in recruiting participants and gathering data, including some methodological challenges and considerations for future work with nonbinary populations. Chapter 4 also contains a detailed account of the creation of the RAINBO corpus. Chapter 5 presents a

brief biography of each participant, including my own subjective impressions of them and some quotes so that they can present themselves "in their own words." While this is a necessarily brief and incomplete snapshot, its intention is to emphasise their status as individuals, with varying motivations and relationships to gender, which I return to throughout the subsequent chapters. Chapter 5 is also meant to provide some account of my positionality as a nonbinary researcher engaging with nonbinary and binary subjects.

Chapters 6 through 9 comprise the acoustic analyses. The approach I have taken here is one that balances breadth and depth, and I investigate nonbinary speech as it manifests within four variables: Pitch and pitch range in Chapter 6, monophthongs and Vowel Space Area (VSA) in Chapter 7, intervocalic /t/ frication in Chapter 8, and every variable in tandem to explore their gendered co-occurrence in Chapter 9. I chose these four variables for a number of reasons:

- ▷ They represent different levels of assumed influence between speaker anatomy and speaker agency.
- ▷ Most of them have been found to exhibit strong sex/gender divisions in previous literature, with relatively little interference from sexuality compared to other acoustic variables such as the quality of /s/ or the aspiration of word-initial /t/.
- ▷ Because they are of particular interest in the context of New Zealand English (NZE), in the case of intervocalic /t/ and monophthongs.

In each chapter, I examine the feature of interest first by establishing a "baseline" assumption in NZE from existing work, if applicable. I outline the predictions born from the three hypotheses, then present the analysis and the results of the analysis. In each case, I discuss some ways the characteristics of the corpus and the incorporation of nonbinary speakers has elicited some methodological considerations or changes in approach. These involve the use of two different kinds of software for handling pitch analysis in Chapter 6, examining both normalised and unnormalised data in Chapter 7, and incorporating multiple scales in reference to binary gender in Chapter 9. In addition to a traditional presentation of the results, each chapter contains a number of sections I have called *Spotlights*, which focus on a the exploration of particular participants or phenomena in the data that are too small to approach statistically, or whose investigation requires a different framework than in the larger analysis. Spotlights are an opportunity to take a qualitative and sometimes ethnographic approach, and reach outside the acoustic domain to incorporate other aspects that shape identity, such as age (Section 7.4), the use of makeup (8.5.3), or familial relationships (8.5). These spotlights serve to contextualise the hypotheses into a broader examination of the complex, multimodal nature of identity construction, affirm the human agency driving the data points in larger analyses, and highlight some potential directions for future research. I conclude in Chapter 10, and summarise the dissertation's contributions and limitations.

Chapter 2

Literature Review, Contextualisation, and Testable Hypotheses

2.1 Introduction

In this section, I provide the background literature review necessary for creating testable sociophonetic hypotheses from the broad research questions presented in the previous chapter. I first examine the scholarship regarding sex/gender in articulatory phonetics and kinematics in Section 2.2. I then examine this same topic in regards to sociocultural lingustics in Section 2.3. Finally, I give an overview of the current literature in sociocultural phonetics examining nonbinary gender in Section 2.4.

2.2 Setting the Stage: Gender/Sex and Articulatory Phonetics

The purpose of this section is to provide some necessary context regarding the way sex/gender is conceptualised and studied within articulatory phonology and phonetics. I maintain (and problematise) a two-pronged approach to variation throughout, examining the limitations to variation imposed both by the physical production limits of human anatomy and the physics of sound, as well as what is ideologically allowed or expected within society.

Generalisation and abstraction are two related ways in which the literature seeks to split (either intentionally or not) the relationship between the physical and social context. Abstraction, via normalisation, seeks to excise physiological variation, leaving only social variation behind. Generalisation, in which the relationship between physical properties and acoustic output is determined in the absence of social considerations, works in the opposite direction. Generalisation is about more than the basic measurement of vocal phenomena, such as modelling the properties of a vocal tract or testing a mathematical formula to predict speaker height from the properties of the voice. Rather, examining generalisation also involves an examination of the outcomes and framing of these measures, where features outside of the prediction space are treated as "abnormal" and those inside it are treated as "normal" and, in some sense, more real. In the same way, normalisation is not inherently about erasing the physiological, but also in examining the ways that abstraction can hide relationships between bodies and speech in misleading ways.

A key challenge in examining this literature is that the physical and the social domains are neither clearly divided, nor do they function independently. Speakers (and listeners) require bodies that interact with the world, and human communication requires a social fabric into which those messages, including the bodies that produce

them, are woven: This complexity makes the presentation of information difficult. In an attempt to organise the information presented in the following sections, I have mimicked the same divisions between sex and gender, biology and society, that I took such pains to problemetise in the previous chapter. While I present the division as illusory, I do not mean to imply that it is completely purposeless. Determining the conditions and extent that people will address tension between the two requires that the popular conception (and limits to that conception) be understood in some isolation. I emphasise it here for a number of reasons: Firstly, because as a structural tool for organising the literature review it breaks a lot of complex and related ideas into more easily-presented "chunks". Secondly, this division of topics represents the direction of the corpus analyses, where the literature tends to split production/perception along speaker/listener lines. As I am primarily examining speaker productions, the question of perception (including self-perception) takes an ancillary role. Finally, because the additional consideration of speakers outside the gender binary, both in this chapter and in the analyses that follow, emphasises that we do not actually make these distinctions in the real world, at least not competently. Even if we are able to separate the hardware from the software, as it were, both are necessary for a functional whole.

This chapter is structured in the following way: In Section 2.2.1, I begin with an examination of the variables of interest (F0, formants, and formant dispersion) as they exist in both nonhuman contexts, as well as in human contexts not related to communicative speech. This includes a discussion of the extent to which anatomy can explain human sexual dimorphism in vocalisation. Even at this first step, the relationship between social expectations and perception plays a role. I then examine generalisation and abstraction in Section 2.2.2. In the last section, 2.2.3, I first extend the acoustic measurements of F0, formants, and formant dispersion to their correlates in speech production and perception; pitch, vowels, and Vowel Space Area (VSA). I then examine two proposed sources for differences in speech across groups, one rooted in easing listener perception - intelligibility; and the other rooted in production efficiency - articulatory settings.

Where applicable throughout the chapter, I also highlight literature specifically concerning Lesbian, Gay, Bi, Trans, and Others (LGBT+) subjects, particularly trans participants. The extent to which I draw from this body of literature, however, is smaller than might be expected given that nonbinary speakers are trans under some definitions (including the one used in this thesis). In the previous chapter, I highlighted some general issues with the way non-cis populations are defined according to different criteria, such as a whole (or partial) rejection of Assigned Sex at Birth (ASAB), or experience of gender dysphoria. This implies a desire to seek medical intervention to mitigate dysphoria along explicitly binary dimensions, and indeed much of the literature studying the speech of trans people is in this framing. This often results in unexpected category divisions of "transgender" participants, such as counting cis crossdressers (Davies and Goldberg, 2006) or anyone who seeks vocal feminization/masculinization (Adler, 2007) as trans, or excluding bisexual people of any sex/gender from the trans category (Kim, 2017).

2.2.1 Sexual Dimorphism: Animal and Human Anatomy

Sexual Dimorphism in Human Body Size

Sexual dimorphism refers to the differences found between females and males of the same species, such as the male lion having a large mane while female lions do not. In

truth, humans have relatively little gross size variation due to sexual dimorphism, at least compared to other animals. In examining height/weight and more granular measures like hip width and mandibular size, variation between men and women within similar groups is found to be around 15% (Plavcan, 2012), though these are highly population-specific (Alarcón, Bastir, and Rosas, 2016, p. 2; Jelenkovic et al., 2016). Conversely, within-sex size variation, the variation found among humans of the same sex, ranges from 10% (for height), to 50% (for weight) (Ruff, 2002). For comparison, the size variation between male and female mandrills is around 340%, with male mandrills being almost three and a half times larger than female mandrills (Setchell et al., 2001, p. 349).

While the overall body size difference in humans is relatively small, other aspects of human physiology that are sexually dimorphic are linked measurably to the acoustic outputs of speech. Changes in puberty cause changes to the internal structure of the vocal system that lower frequency in men without increasing overall size. However, physiology alone is often not enough to fully explain the acoustic differences, nor the gendered perception of voices.

F0

The fundamental frequency, F0, is linked to the properties of the vocal folds themselves - thickness, length, elasticity, etc. F0 is a measurement in Hz of the rate at which a specific set of vocal folds vibrates as a product of the aforementioned properties, while pitch is the way F0 is perceived or interpreted by human listeners (Li and Jain, 2009). Outside of this chapter, I use F0 and pitch relatively interchangeably, but tend to use "F0" to refer to specific Hz measurements and "pitch" when I am presenting F0 relative (higher/lower) to other speakers or measures.

While the relationship between F0 and pitch is one of production vs perception, both are also subject to the influence of empirical and social limitations: F0 is controlled not only by immutable physical properties such as vocal fold length or elasticity, but by the rate of airflow and degree of muscular tension present in the vocal folds, which are controlled by the speaker and subsequently subject to social conditioning (Davis et al. (1996), Portela, Hammarberg, and Södersten (2013), and Davidson (2021), among many others). Pitch perception, in addition to the social expectations involved in associating certain voices with certain speakers, is also subject to the limitations imposed by the central auditory nervous system in identifying and differentiating different pitches (Tramo, Shah, and Braida, 2002).

The correlation between larger overall body size and lower pitch, known as negative size-frequency allometry, has been confirmed across many species, including songbirds (Linhart and Fuchs, 2015), frogs (Micancin and Wiley, 2014), carnivores and dozens of non-human primates (Bowling et al., 2017; Rendall et al., 2005), such as the previously-mentioned mandrills. In humans, however, most studies have found weak or no evidence of negative allometry between F0 and speaker size, as measured either generally across populations when age and sex/gender have been controlled for (Künzel, 1989; González, 2004). Pisanski et al. (2016), however, found F0 to significantly predict around 0.5-2% of the *within-sex* variance in height/weight. The results of these studies seems at odds with the logical connections we can draw between humans and the animal kingdom: If larger animals produce lower pitch, and men are larger than women (even if that difference is only around 15%), men should, therefore, produce proportionately lower pitch than women. In fact, men do produce lower pitch, but not in a simple 1:1 relationship with body size: In a sample of 68 Canadian undergraduate students (34 binary men and 34 binary women), Rendall

et al. (2005, pp. 947–948) found that the difference in mean voice pitch between males and females greatly exceeded any difference in body size between the sexes.

During puberty (and/or in hormone therapy), greater levels of testosterone cause the pharynx to lengthen and the vocal folds to thicken, resulting in physiological differences in pitch between men and women beyond that accounted for by body size alone. These pitch differences have been proposed by Fitch and Giedd (1999), Rendall et al. (2005), and Evans, Neave, and Wakelin (2006), among others, to be a secondary sex characteristic that developed before speech in humans, the function of which is similar to a bison's hump, to make men seem larger than they are. This is corroborated by Van Dommelen and Moxness (1995), who found that, when presenting speech samples of male and female voices to both men and women, only male listeners were able to accurately discern male speaker size from male voice samples, and lower pitch elicited larger size estimations among listeners regardless of actual speaker size. Both Van Dommelen and Moxness (1995) and a separate study by Armstrong, Lee, and Feinberg (2019) likewise found strong evidence of negative size-frequency allometry in the *perception* of male voices: That is, people attribute a much greater relationship between body size and pitch than actually exists, when estimating men's heights from voice samples. While I speak more about gendered differences in production later, this particular example highlights the difficulty in trying to isolate the physiological influences on production from social ones: While a size difference does exist between the sexes, its influence is disproportionate on both production and perception, in opposite directions.

There has long been a distinction made between overall Vocal Tract Length (VTL) (responsible for lower formant dispersion, discussed in the next section) and vocal *fold* length, responsible for lowering the fundamental frequency specifically (Fitch, 1997, p. 1519). This is confirmed by Zhang (2016, p. 1506) who, using a computational model of the human vocal tract, found that F0 increased as vocal fold thickness decreased, and F0 also increased as anterior-posterior stiffness in the vocal folds increased. This paints a slightly more complicated picture of the relationship between F0 and the physical properties of the vocal tract, given that androgens are implicated in both the thickening *and* stiffening of vocal folds but do not increase physical height or size on their own (Papp, 2011, p. 37; Hunter, Tanner, and Smith, 2011, p. 3). To sum up, while there is very little relationship between F0 and general body size in humans, sexual dimorphism of the voice exists between the sexes as a product of changes inside the vocal tract during puberty, and this results in a very strong *perceptual* relationship between pitch and size in our interpretation of human voices.

Formants & Formant Dispersion

Formants above F0 are the resonant frequencies caused by the interaction of the vibration of the vocal cords with the rest of the vocal tract. They are numbered according to their distance from F0: The closest resonant frequency above F0 is F1, the next is F2, etc. The default frequencies of formants, and the distances between them (known as formant dispersion), are determined by the overall length of the vocal tract and the relative sizes of structures within it. Formants (particular F1-F3) can be altered independently via the manipulation of structures inside the vocal tract, such as the lips and tongue, and the acoustic output of these formants corresponds to the linguistic production of vowels. The output of overall formant dispersion corresponds to VSA. I give a more detailed explanation of the way both formants and VSA are measured and interpreted in Chapter 7.

Because of the uniqueness of human speech in manipulating formants for communication, animal studies examining formant frequencies specifically are quite rare, especially outside of nonhuman primates. Within primates, formant dispersion has been found to correlate with body and muzzle size in baboons (Rendall et al., 2005) and in distinguishing across similarly-sized species in lemurs (Gamba, Friard, and Giacoma, 2012). While ten (or more) formants can be found in human speech sounds, they are not all of equal importance to communication in English: The first two formants (F1/F2) and, to a lesser extent, the third formant (F3) are of primary importance in the interpretation of vowel sounds in English. F1 through F4 have been found to correlate with speaker height, though this is highly dependant on the vowel used, and which formant or combination of formants is measured (Greisbach, 2007).

Vocal tract length (VTL) is correlated to formant frequencies in an inverse relationship: the longer the vocal tract, the lower the formants (Stevens, Kasowski, and Fant, 1953; Chiba and Kajiyama, 1958; Childers and Wu, 1991; Arai, 2004). Lammert and Narayanan (2015) recently confirmed this finding using modern equipment via the comparison of a computer model of the vocal tract with human measurements via MRI, and were able to predict VTL to within 8.4 mm, given enough acoustic data from the upper formants. Johnson (2018a) found that, overall, vocal structure anatomy predicts between 26-36% of variance between speech segments (vowels and fricatives). Overall, the data indicates that the relationship between sex/gender and VTL is much stronger than the predictive relationship between sex/gender and F0,[1] but still much lower than we would expect if formant production was purely driven by asocial factors like anatomy.

In addition to the differences found in formant frequencies above, formant dispersion in nonhuman animals is found to correlate with general body size among many species including koalas (Charlton et al., 2011), baboons (Rendall et al., 2005), and macaques (Fitch, 1997). In humans, González (2004) finds a weak correlation between vowel dispersion and body size (height and weight), though the correlation is stronger for women than men. Like F0, human formant dispersion in production exhibits stronger sexual dimorphism than differences in body size alone would suggest. Overall, men tend to have more tightly stacked formants, while women's formants are more widely spaced. Physiologically, this is proposed to be caused by post-pubital men having relatively longer pharynges (and thus overall greater VTL) than women, though Fitch and Giedd (1999, p. 1514) find a large degree of overlap in across-sex VTL. According to Armstrong, Lee, and Feinberg (2019), "About 75% of the explanatory power in height is lost when VTL is estimated from formant frequencies as opposed to being measured in MRI." (p.44).

Conclusion

While it is clear that differences in physiology across binary sexes have some acoustic consequence when it comes to F0 and formants, it is equally clear that these consequences have limitations. They are neither exclusively gendered/sexed (as variation between sexes and across populations is wider than that found between sexes of similar populations), nor the entire story (as evidenced by the relatively small amount of variance, 0.5-36%, that they are found to explain).

[1] Weirich and Simpson (2013) found no significant correlation *between* Pitch and VTL (via VSA)

2.2.2 Generalisation & Abstraction

This section highlights two specific subsets of literature that relate to the relationship between the physical and social influences on speech. The first explores the way that early scholarship on measuring the human vocal tract has had some features (and complications) generalised over time, which can serve to conceal points of variation and place disproportionate emphasis on "defaults." This generalisation, in turn, can influence the social perception of what is normal or expected physiological variation. The second topic is an examination of the deliberate methodological abstraction of scaling and normalising across individuals, with the goal of erasing physiological differences so that only social variation remains. This approach is problematic as it presents speech as something that can be extracted completely from the body that produced it.

Generalisation: The Male Vocal Tract as the Default[2]

This section serves to briefly highlight a relatively small, but far-reaching, tendency in the articulatory phonetics literature to extend the interpretations of primary data to the point where the extension becomes axiomatic. Consider the following two excerpts, the sources of which boast over nine thousand citations combined on Google scholar:

> "From a material collected by the radiologist Paul Edholm of Karolinska Sjukhuset, Stockholm, the details of which have not been published I have collected the following measurements. A male articulating the vowel [i] had a pharynx length from glottis to the soft palate of 9.1 cm, and a mouth length from the incisors to the back pharynx wall of 8.25 cm. The corresponding female measures were 7.0 cm and 7.0 cm respectively."
>
> Fant, 1966, p. 27.

> "Since the average length of the vocal tract of males is about 17.5cm, the resonances appear at [...]"
>
> Jakobson, Fant, and Halle, 1963, p. 18.

The cadaver measurements referenced in the Fant (1966) quote above are from a single male and single female Swedish speaker. The data in Jakobson, Fant, and Halle (1963) was, I believe, obtained/extrapolated from x-ray data on an unknown number of English/Swedish speakers measured by Fant himself.[3]

By modern standards, it is somewhat problematic to present this small subset of speakers as representing the default measures for men and women, though it was common practice at this point in history. However, a more egregious issue arises over time: skip ahead a half-century, and the male data is now presented in multiple works as the *human* default. Some examples follow (I have added emphasis);

> "For convenient comparison with **typical adult values**, the formants were computed as if the infant's vocal tract were **17.5cm** long."
>
> Lacerda, 2003.

[2] This section would not be possible without the contribution of Julie Abbou. Her talk at Sociolinguistic Symposium 22 (Abbou et al., 2018), and the conversation afterwords, introduced me to a rich body of French feminist scholars and scholarship, including Arnold (2016), which first introduced the problematic use of Fant's reported lengths measures, and from which I take some of the quoted examples.

[3] The work cites a number of quarterly and technical reports from the M.I.T. Acoustics Laboratory (Fant, 1951; Fant, 1950a; Fant, 1950b; Fant, 1952), but they are scant on details.

> "Thus, with an average vocal tract length about half the size (8 cm) of **an adult human (17 cm)**, an infant [...]"
>
> Boë et al., 2007, p. 571.

> "In the normalised measurement, all vocal tracts were scaled to a **mean length of 17.5 cm**[...]"
>
> De Boer, 2010.

> "[...] as betrayed by his resonance frequencies being noticeably lower than those produced by **a normal vocal tract of 170-mm** length [...]"
>
> Schroeder, 2013, p. 6.

A second, related issue from the same source regards the relative difference in vocal tract size between men and women. The relevant section of the source material outlines a general scale factor (18%, similar to the general body-size difference of 15%), as well as some deviations.

> "The female to male scale factor is of the order of 18% averaged over the whole vowel system. The typical deviations from this rule are:
>
> (a) The first and second formants of rounded back vowels have relatively low scale factors;
>
> (b) This is also the case with the first formant scale factor of any close or highly rounded vowel, i.e. high front vowels;
>
> (c) Very open front or back vowels display a first formant "Sex factor" k_1 which is substantially higher than average."
>
> Fant, 1966, p. 29.

In addition to the cadaver measurements from two Swedish speakers above, this scaling factor incorporates acoustic measurements Fant obtained himself from 7 male and 7 female Swedish speakers, and the 76 American English speakers (15 children, 33 men, 28 women) from Peterson and Barney (1952). As we move through time this measure becomes simplified to 15%, and likewise axiomatic:

> "Supralarygeal vocal-tract length (VTL) is **about 15%** greater in males (Fant, 1966), producing characteristically lower formant frequencies."
>
> Bachorowski and Owren, 1999, p. 1055.

> "Because of the lowering of the male larynx in puberty, adult males' vocal tract length (VTL) exceeds females' VTL by **about 15%** (Fant, 1966)."
>
> Skuk and Schweinberger, 2014, p. 285.

> "Because the male vocal tract is **about 15%** longer than the female vocal tract, the speech of men can be expected to have lower formant frequencies than those considered characteristic of women."
>
> Gelfer and Mikos, 2005, p. 545.[4]

In both of these cases, following the citations often leads to either one or both of Fant (1966) or Jakobson, Fant, and Halle (1963). I do not want to over-exaggerate the issue as systemic, though it is far-reaching: In literature focussed on more applied fields like communication disorders and speech language pathology, you are more

[4] This paper does not cite Fant directly, but cites the first quote, Bachorowski and Owren (1999).

likely to see Baken and Orlikoff (2000) cited for male and female averages, which contains data from multiple clinical measurements of the voice. Baken and Orlikoff (2000) itself is not completely free of a "male-as-default" bias (reporting on a total of 369 male subjects in spontaneous speech compared to 6 for women, for example). Often, papers will cite both, using the pitch mean and range data from Baken and Orlikoff (2000, pp. 173–177) and the VTL data from Fant (1966).

I present these two examples to draw attention to the underlying assumptions rather than to try to pick apart the conditions under which the information was gathered. In this framing, however unintentionally, some half of the population (binary women) are considered a deviation from the "norm." This is far from a new or recently-discovered issue, even within an exclusively binary framework: McConnell-Ginet (1978) and Henton (1989) both address it directly, pulling no punches in describing and denouncing what the first refers to as androcentric biases[5] and the second as "overtly misogynistic characterizations of female speech" (Henton, 1989, p. 299). From a more contemporary perspective, Arnold (2016) contains a much more detailed analysis of the clinical and social ramifications of this scientific othering than I could provide here, but in summary: Within the structure presumed and perpetuated by this literature, women are not only excluded from normalcy, but they make up the *entire* exclusion: Given a binary system, if you presume one of the two options is the default, by definition anything not in the default category must make up the "other" category. Anything not-male is both abnormal and *de facto* female. For non-binary speakers, the influence of patriarchy adds further complexity to an already complex task: Not only must they navigate stereotypes regarding binary gender, but they must do so with the knowledge that those genders do not exist in a perfectly balanced vacuum. Rather than being a problem specific to phonetics, this bias in reporting is a reflection of the society in which it takes place, in which bodies are categorized into binary sex, based on some set of assumptions (which I have shown in the previous chapter to be diverse and ideologically-motivated) that presumes homogeneity within gender groups beyond the categorization criteria. I discuss these broader social conceptions in more detail in Section 2.3.2.

Abstraction: Normalisation

The literature on gender differences in speech is closely linked to the literature on speaker normalisation. The process of normalisation, in its most general sense, seeks to align measurements from disparate groups along a common scale. Once this is done, you can determine the amount of similarity or deviance each speaker or group displays from the collective norm. In practice, in linguistics, this usually takes the form of normalising across all speakers in a given group (or all groups of data, such as vowels, from a given speaker). The result is a process of classifying systematic variation as physiological and separate from the socially-driven variation that applies to smaller units within that system. This is intended to mimic the process implemented to varying degrees by listeners to separate the "extraneous" physiological information that differentiates talkers from the more "relevant" linguistic signals differentiating speech (Peterson, 1951; Johnson and Sjerps, 2021). While a full literature review of normalisation in speech *perception* is beyond the scope of this section, the connection between them is worth illuminating: Barreda (2020, p. 2) makes the connection between speech perception and production explicit, proposing that vowel normalisation

[5] "The fact remains, however, that women's tunes will be interpreted and evaluated from an androcentric perspective." (McConnell-Ginet, 1978, p. 542). The use of *tunes* here refers primarily to phrase-level intonation, but also pitch, pitch range, and pitch dynamism.

in perception is a form of perceptual constancy, where what listeners know (or think they know) about the speaker and the world shapes their interpretation of the signal.

Normalisation techniques can be sorted into those that treat certain key factors, such as speaker, vowel, and formant, as either intrinsic or extrinsic (Flynn, 2011, p. 3). Intrinsic procedures are those which can be measured within a single category member, without reference to others of its category, while extrinsic techniques regard the entire system. The goals of any given normalisation technique can be sorted according to its prioritisation of four broad goals: (1) Eliminating physiological variation across speakers; (2) Preserving variation caused by non-physiological (social) categories; (3) Maintaining linguistically significant differences across categories and phonemes, and; (4) Modelling the cognitive process of normalisation employed by human minds in interpreting speech (Flynn, 2011, p. 2). I am concerned here with (1) and (2) – eliminating physiological variation while preserving social variation. The relationship between those two sources of variation once again creates an inherent tension between goals that seek to prioritize one over the other.

The forces driving "physiological" and "social" variation are rooted in a set of properties that create the illusion of total distinction, but the areas of influence and overlap of each on the other makes them inextricable at a base level. Therefore, any process (such as normalisation) that seeks to completely separate the influence of one over the other is doomed to under- or over-specification. Flynn addresses this issue obliquely by acknowledging that all normalisation has drawbacks, and that the concerns of the particular research question should be the driving force as to *which* subset of drawbacks are chosen. Johnson (2018b) also points out that, even after the "best" (most ideal for the dataset) form of normalisation is chosen, the degree to which any process should be applied is also a matter of researcher agency: "Knowing when a method “over-normalizes” and removes talker variability that actually “should” remain because it is sociolinguistically significant, is a huge problem." (p.7). Zimman (2021, p. 73) summarises this major issue with normalisation with regards to sex/gender specifically: "Importantly, constructivist approaches to gender do not maintain that the body is irrelevant. Instead, 'biology' is seen as always filtered through some sort of cultural lens and never completely separated from social practices." Any acoustic criteria which seeks to define physiological (sex) limitations as fully distinct and distinguishable from social (gender) limitations is not only going to fail to categorise *all* nonbinary people correctly, but is also going to incorrectly categorise a subset of binary speakers (both trans and cis) via over- or under-normalisation.

The first large-scale study identifying broad acoustic differences in vowel production across groups of American English Speakers (men, women, and children), and how to effectively normalise across them, is outlined in Peterson (1951) and Peterson and Barney (1952). While the normalisation procedure proposed in that work has become one of many, the production measurements in Peterson and Barney (1952) (and the later replication/extension in Hillenbrand et al., 1994) have become a "gold standard" dataset for testing normalisation techniques in phonetics. While a consideration of the history and current conversations in vowel normalisation is beyond the scope of this thesis, I recommend Flynn (2011), Johnson (2018b), and Brand et al. (2019) for detailed comparisons of different approaches, as well as theoretical and practical examinations of normalisation as it refers to formants, the vocal tract, and the whole vowel system, respectively. I address the specific techniques used in this thesis in Chapters 7 and 8 where I utilise normalisation in the analysis.

Though many methods of normalisation are convenient and a relatively simple process to apply, it is impossible to avoid a certain amount of researcher-influenced

bias in determining:

1. Which groups are appropriate to normalise across.
2. What kinds of misclassifications are acceptable.
3. Which individuals, and which data points from those individuals are relevant, and which are outliers.
4. Which categories are considered relevant, which category information to collect in the first place, and, if applying speaker-intrinsic normalisation processes, how to regroup speakers into sensible categories after normalisation.

The last point is arguably not intrinsic to normalisation, which itself, taken as a bare mathematical process, is agnostic to the group members fed to it. However, I present it as problematic here due to the way normalisation generally functions in practice, as one step within a larger research paradigm. To illustrate with an example, assume a population that contains both binary trans and binary cis speakers, as well as nonbinary trans speakers. One can imagine different justifications for categorising binary trans men, post-menopausal cis women, or nonbinary intersex people into different groups, based on the inclusion/exclusion criteria. While group membership is irrelevant in that the same speaker-intrinsic normalisation technique can be applied to every speaker, the interpretibility of that data relies on the pre- and post-categorisation. If a researcher intending to normalise collects data on speaker gender but not ASAB, or ASAB but not hormone profile, they have already made an assumption about the relevance of those features to an analysis.

In the literature reviewed here, despite the data rarely falling neatly into sex/gender categories as proposed, the fundamental suitability of "sex" as a category is rarely questioned, and even speaker-intrinsic normalisation methods cease to be truly intrinsic in application once speakers are binned back into predetermined categories after normalisation. The inflexibility of baseline assumptions regarding sex as binary/static and gender as enacted/mutable imposes an artificial dichotomy between the physiological and the social sources of variation. Many speakers, not just nonbinary speakers, simply do not "fit" anywhere within this current theory in variationist linguistics, without addressing these basic assumptions regarding sex/gender. Fortunately, the literature itself provides its own solution: In virtually all cases examined here, individual variation within gender groups outweighs gender difference across groups. Throughout the broader linguistics literature on gendered discourses are examples of (cis binary) individuals whose enactment of their own sex/gender challenges these binary axioms, often including other aspects of their social and physical identity, such as race, class, ability, and sexuality. These examinations cover both binary women's relationship to femininity (Sa'ar, 2007; Holmes and Schnurr, 2006; Kendrick, 2008), and men's relationship to masculinity (Myketiak, 2016; Isaacs and Swartz, 2022). The strength of the axiom is further illustrated by the lack of consideration for a nuanced relationship between a binary person's sex/gender and the hegemonic stereotypes of the "other" sex/gender, outside of examinations of sexuality. That is, the relationship between cis binary women's relationship to hegemonic masculinity and cis binary men's relationship to hegemonic femininity, is often ignored or framed as oppositional. This is often extended or reinterpreted in examinations of sexuality, such as Livia (2005, p. 252) where the butch/femme stereotype is presented as, simultaneously, a reenactment of binary sex/gender division, a radical critique of that division, and a reinterpretation of femininity to include multiple ways of enacting womanhood.

This indicates that sex/gender is unlikely to be the consolidated bedrock upon which other variation is layered, particularly in consideration of the growing dissatisfaction with the binary among both LGBT+ and cis populations identified in Chapter 1. It is possible to acknowledge and account for the influence of physiology and the social expectations that arise from it, while also questioning and iterating upon the use of research paradigms that create a false division between the physical and the social. Acoustic studies with trans and other gender nonconforming participants already provide some useful templates for speech research in all populations, such as splitting up gender from ASAB and sexuality (Clark et al., 2018), asking about or measuring hormone balances (T'Sjoen et al., 2006; Zimman, 2017), and/or explicitly examining the vocal tract via tools like laryngoscopy (Cler et al., 2020). In many of these cases, normalisation is still a useful and relevant tool, when utilised within and regulated by a more deliberate and reflexive construction of the surrounding research methodology.

2.2.3 Gendered Speech: Human Production and Perception

In the previous sections, I highlighted the limitations of examining the relationship between human bodies and voices from a purely allometric perspective, and the way that approach has shaped the literature via generalisation and abstraction. Having outlined the exploration of what is possible (or expected), I turn now towards what we actually find in speech production and perception. This section follows the same structure as Section 2.2.1, but with a focus on what we actually find in human speech, rather than on what sounds are physiologically possible to produce. F0 is first expanded into pitch[6] and pitch range, then formant production and dispersion serve as the jumping-off point for an examination of vowel production and VSA. Finally, I examine the relationship between production and perception and speaker gender as it relates to intelligibility and articulatory settings.

Pitch and Pitch Range

While speaker gender predicts significant differences in F0, individuals exhibit wide variation within-group, and same-gender speakers of different languages do not present equivalent differences in F0 (Takefuta, Jancosek, and Brunt, 1972; Pépiot, 2015). It should be noted that these results are subject to the same "male as default" approach I critique in Section 2.2.2, resulting in the values reported here for men and women being based on unbalanced populations, with the number of women making up representative samples being a fraction of those for men. Linke (1973) points out this lack of data, and the resulting poor agreement across studies on women's speech, in the 1970s. Thirty years later, the issue persists, as Baken and Orlikoff (2000, pp. 173–177) aggregate the results of multiple studies and find spontaneous speech measures from 369 men, compared to 6 women. It is my impression that this gap has not closed much in the intervening 20 years, with contemporary papers using the same, or older, data when reporting averages for men and women (for example,

[6]The literature focussed on clinical practice or vocal therapy often frames the distinction between F0 and pitch as one between *habitual* and *optimal* (or natural) pitch (Britto and Doyle, 1990). While habitual pitch is the frequency more often produced by an individual, optimal pitch is the one produced with the most efficient use of resources, given a person's anatomy. This makes sense if your patient's goal is to reduce the effort involved in speech, via reducing the disparity between the two. The value judgement inherent in the naming of the idealised pitch as natural/optimal was pointed out as problematic in every paper that used it. However, one of the participants in this research, Istus, uses "natural" to refer to her voice in this way, so it is a relevant description for some speakers.

Davidson, 2019; Pépiot, 2015). Research using this unbalanced data indicates that, in production, binary women display wider variation in mean pitch than binary men: Measured means for cis men in the papers cited in this section range from 100Hz to 136Hz (Baken and Orlikoff, 2000, pp. 173–177; Davies and Goldberg, 2006, p. 178), while for women they range from 185 to 318Hz (Linke, 1973; Davies and Goldberg, 2006; Takefuta, Jancosek, and Brunt, 1972).

Perceptually, pitch tends to be the most salient predictor used by listeners in determining speaker sex in English (Wolfe et al., 1990; Coleman, 1976; Whiteside, 1998; Bishop and Keating, 2012). Unsurprisingly, it is also the primary target in terms of vocal therapies for speakers that wish to shift listener perception of their gender. According to an overview by Gelfer and Mikos (2005, p. 545), multiple studies "have shown unequivocally that increasing speaker fundamental frequency is of primary importance when trying to shift perception of the voice from man to woman."[7] Ryalls and Lieberman (1982) and Bishop and Keating (2012) identify the area around 200Hz as the perceptual boundary in determining speaker gender: In both natural and synthesised speech, listeners' ability to correctly identify binary voices drops drastically above this frequency when identifying men, and below this frequency when identifying women. There is also an area of overlap in men's and women's voices between 145Hz and 165Hz (Oates and Dacakis, 1997, p. 178) and sometimes as high as 181hz (Gelfer and Schofield, 2000), that is often categorised as "gender-neutral". However, pitch is not an *exclusive* cue to the perception of gender: Trans women speaking in this "gender neutral" range are often perceived as male (Davies and Goldberg, 2006), and multiple studies have found that when F0 is deliberately masked or confused, listeners use other signals to determine speaker gender, such as pitch range, intonation, and formant dispersion (Childers and Wu, 1991; Murry and Singh, 1980; Davies and Goldberg, 2006; Bishop and Keating, 2012).

Vowels

As explored in Section 2.2.1, while the higher formants (F4 and up) tend to carry indexical information about speakers. F1 and F2 are used primarily in English to distinguish vowel sounds in speech[8], and therefore are the primary focus of this work. Women have been found to produce higher formants than men (Henton, 1995; Whiteside, 2001; Pépiot, 2015), but this variation, while consistent, does not manifest uniformly across languages, across the vowel inventories within a single language, nor even across F1 and F2 within a single vowel's production space. Both F1 and F2 of vowels that representing extreme tongue positions (like FLEECE, GOOSE, and TRAP) are found to vary more as a consequence of speaker gender than vowels produced more centrally. Weirich and Simpson (2013, p. 2966) summarises this tendency clearly: "In general terms, the higher the formant value the greater the difference between male and female values." Across all vowels, F2 is found to vary more than F1 as a consequence of speaker gender. In more specific terms, this results in differences in the F2 of FLEECE to be the most consistent marker of gender variation in multiple dialects of English (Pépiot, 2009; Weirich and Simpson, 2013; Coleman, 1976).

[7] Gelfer and Mikos (2005) also found that listener perception of the voices of trans women patterns most closely with cis women in all cases, though it should be noted that their methodology also excluded trans (but not cis) women whose F0 during "sustained vowel production" (pg.548) dropped below 165Hz at any time.

[8] F3 also varies consistently with vowel sounds, particularly lip-rounding.

Perceptually, it is also the case that people use F2 as a marker of sex/gender. Listeners are better at successfully identifying the gender of a speaker using F2 compared to other formants, both when formants are presented in isolation (Childers and Wu, 1991), as well as in cases where the whole signal has been manipulated to lie between naturally recorded productions from men and women (Skuk and Schweinberger, 2014). These differences, found within vowel production and perception, and even within formants of a single vowel, are indicative of the position of speaker sex/gender as a social factor in the larger system of speech production. Indeed, it is the case that women are traditionally implicated as leaders in sound change (Labov, 1994), with caveats: Age, social class, and the stigmatisation/prestige of the variable in question all influence production, in that younger, lower-class women are said to lead non-stigmatised changes, and that all other women are somewhat more conservative in production (Maclagan, Gordon, and Lewis, 1999). More recent work has highlighted further nuance in these results, such as the finding that individuals of either binary gender may be more likely to "lead" or "lag" in reference to sound change (Brand et al., forthcoming). Other work, beginning with Eckert (1989) and others which I explore in more detail in Section 2.3.3, has gone farther, and in takes issue with the use of sex/gender and other macrosocial categories as explanatory paradigms, as a source of variation in itself.

Vowel Space Area (VSA)

The correlate to formant dispersion discussed in Section 2.2.1 is VSA. While formant dispersion is a measure of the distance between formants in non-speech vocalisation, in speech sounds the location of and relative distance between specific formants matters: In English, F1 and F2 serve to separate particular vowels from each other. Overall formant dispersion is dictated by VTL and the relative position of the larynx, while specific formant position in vowel production is a consequence of the interaction between anatomy and speaker agency, hence why examinations of the latter are rare in non-human animal studies. For example, a fallow buck can drop its larynx to reduce formant dispersion (McElligott, Birrer, and Vannoni, 2006), but it has no control over the particular or relative position of those formants: It is vocalising, its formant dispersion is communicating something to other bucks in the area, but it is not producing vowels and is not *speaking*.

When examining the use of formants to produce different vowels in speech, it is common to present the relationship between F1 and F2 in vocalic sounds as a point on an abstracted vowel space. If we conceive of this abstract acoustic space consisting of F1 on one axis and F2 on the other, we can interpret a single production of a single vowel as existing at a single point in that abstract space[9], and the production space of a single vowel as the area inside which most productions (points) of that vowel fall. Broadly, VSA refers to the measurable limits of that abstracted acoustic space, the area that contains all of a given speaker's productions of every vowel. The absolute possible limits of the VSA are defined by a speaker's anatomy, but the actual VSA is smaller than that, as it consists of measured productions (sometimes with outliers removed). I define the ways I measure and analyze the VSA of speakers in the RAINBO corpus more explicitly in Section 7.1.1. By convention, the vowel space is visualised as a trapezoid with F2 along the X-axis and F1 along the y-axis. This allows the abstract space to correlate roughly with human vocal anatomy, where the back of the tongue's relative height aligns roughly with measurements for F1

[9]Or as a line connecting a start point and an end point, in the case of diphthongs.

(where a low F1 correlates to a high tongue body, and a high F1 correlates to a low tongue body), and the relative saggital position corresponds to measurements for F2 (where a high F2 is regarded as a fronted vowel with the tongue peak closer to the teeth/lips, and a low F2 is regarded as a backed vowel with the tongue peak closer to the throat)(Ladefoged and Johnson, 2014, p. 22).

Neel (2008) found VSA strongly correlated with formant dispersion, and most strongly in the F2 dimension. However, the limited relationship between size (in terms of of both the length of the vocal tract and the general size of the body) and formant dispersion, as explored in Section 2.2.1, is not enough to explain the sex/gender dimorphism in VSA found in human speech. Goldstein (1980) developed an articulatory model of the vocal tract that incorporated growth parameters to examine changes in the vocal tract from infancy through adulthood (including developmental differences between men and women). She compared the output of this model with the expected vowel measurements from Fant (1966)'s real-world data and found that, once anatomical differences in VTL are accounted for, women's vowel spaces were wider for all vowels when compared to men.

Supporting Goldstein's theoretical models, many studies find binary women to have wider vowel spaces than binary men, across languages, both after physiological differences are normalised away (Fant, 1966; Fant, 1975; Henton, 1992; Diehl et al., 1996; Yang, 1990; Munson, 2007)[10], and in cases where little-to-no physiological differences exist, such as before puberty (Whiteside, 2001; Pettinato et al., 2016; Vorperian and Kent, 2007). However, other studies have found no relationship between gender and VSA after formant normalisation when examining speakers from multiple dialect areas (Jacewicz, Fox, and Salmons, 2007).

These results offer robust, if complicated, evidence that sex/gender differences in language include a strong behavioral component, but the driving force behind the behavior is unproven. Multiple potential motivations have been proposed for these differences in production, two of which I examine here: intelligibility, whose source is the desire for speakers to make themselves understood as compensation for the limitations of anatomy; and articulatory settings, whose source is the most kinematically efficient way to speak, given the structure of the vocal tract and sound system of a given language/dialect/register. While both incorporate social and physiological drives in production, the primary influence varies: Examinations of intelligibility tend to emphasize the former, as a social desire to be understood resulting in changes to speech to reduce the influence of physiology. Articulatory settings, conversely, emphasize the latter, as rooted in a drive to conserve mechanical energy.

Intelligibility

The intelligibility of speech, generally measured in experimental conditions by the ability of listeners to perceive speech (or speech sounds) in noise, has often been linked to the relative size of the vocal tract. While Fant (1975), when discussing normalisation, generally focusses on *how* to account for the production differences found between men and women, he does suggest, in passing, two possible reasons *why* these differences are non-uniform and unaccounted for by physiological factors

[10]Henton (1992) and a later conference proceeding by the same author (Henton, 1995) are often cited, but I have been unable to find more than an abstract for either work. I am not sure what specific languages, normalisation procedures, or conclusions Henton offers besides what is paraphrased in other works (that women's spaces are wider in seven different languages and three English dialects). I hesitate to include it, but it is cited often enough that it feels remiss to leave it out.

alone. The first, later known as the compensatory hypothesis, is rooted in a desire for clarity: Women's higher pitch results in less perceptual distinction between different sounds, and so women must compensate for their smaller vocal cavities via greater constriction, "to satisfy demands of perceptual invariance or contrast" (pg. 17). The second is a "universal "feministic" preference in vowel qualities" (p.18), which might be interpreted as type of articulatory setting, which I explore in the next section. Goldstein, like Fant, attributes these differences to the adage that women desire to speak more clearly/intelligibly, at least in Western languages (Goldstein, 1980, p.233). Indeed, women's voices are generally ranked as more intelligible than men's voices in English overall (Bradlow, Torretta, and Pisoni, 1996; Kwon, 2010).

The hypothesis that intelligibility is tied to VSA is partially supported experimentally by Diehl et al. (1996), who found that, independent of gender, speakers with wider vowel spaces tend to be ranked as more intelligible by listeners. However, their findings also challenge the compensatory hypothesis, as they did not find a linear correlation between F0 and intelligibility. Incorporating an examination of F0 into the link between VSA and intelligibility, Simpson (2011) found significant correlations between F0 and VSA in both men and women after first identifying it in women only in Simpson and Ericsdotter (2007), though in the latter case only *after* removing speakers from both binary gender groups (men and women) with the highest and lowest overall VSA, resulting in the removal of 15% of the total sample. Later work by one of the same authors found no significant correlations, and attributed the earlier results as likely due to differences in sampling – that is, in treating "extreme" F0 values as normal (Weirich and Simpson, 2013, p. 2972). To sum up:

1. Women's voices are generally judged by listeners as more intelligible.
2. There is (sometimes) a linear correlation between F0 and VSA, in that a speaker with a higher F0 will have a wider VSA. This correlation is more likely to be found among women than men.
3. There is a positive correlation between VSA and intelligibility, regardless of gender.
4. There is no linear correlation between F0 and intelligibility.

This indicates that vowel space, F0, or the combination of the two are not the sole reason for intelligibility scores, and that listener assumptions about women's voices and intelligibility also plays a role. While a full accounting of this literature is beyond the scope of this chapter, Diehl et al.'s findings are typical in that, regardless of speaker gender, "atypical" speech – that is, speech that does not align with *listener expectations* – is regarded as less intelligible (Ryalls and Lieberman, 1982; Bradlow, Torretta, and Pisoni, 1996; Diehl et al., 1996). The complex interaction of intelligibility in production and listener expectation is an important factor in the examination of non-binary speech (since nonbinary identities themselves are unexpected for most listeners), and one I return to throughout this thesis. But how does a listener know what to expect, given the link between expected production and measurable physical features (such as size) cannot be established, and the link between production and secondary sex characteristics is not complete? Articulatory settings may offer one possible explanation.

Articulatory Settings

Conceptually, the idea that different languages require different overall vocal postures for native-like speech has been proposed by philologists and phoneticians since

the mid 17th century (for a detailed historical overview, see Laver 1978). A more formalised definition of this concept under the term *articulatory settings* was described by Honikman (1964), using examples from teaching second-language speakers to employ a host of complex-yet-subtle changes to speak with less of an accent. The articulatory settings of a language, as described by Honikman, refers to "the gross oral posture and mechanics, both internal and external, requisite as a framework for the comfortable, economic, and fluent merging and integrating of the isolated sounds into that harmonious, cognizable whole which constitutes the pronunciation of a language." (p. 73). Laver and Trudgill (1979, p. 14) refer to the same phenomenon in the context of dialectal differences, as "the way in which the speaker 'sets' his muscular vocal apparatus for speaking [...] habitual[11] tendencies towards maintaining particular muscular adjustments." They use the term *voice settings* here rather than articulatory settings, as the latter term was generally reserved, at the time, as referring only to cross-language differences in articulatory tendencies (Laver and Trudgill, 1979, p. 15).

From Laver's work in the 80s, the literature splits into two distinct, complementary streams. The first, following more directly from the examination of voice settings, generally adopts the term "voice quality" and, as that name suggests, is less concerned with the specific articulators and focuses instead on the acoustic output. The link between the voice quality and affect, stance, and other identity markers takes precedence to the link between the voice and the body that produces it. As (Podesva and Callier, 2015, p. 187) remarks "Treating the voice as inalienable from its speaker is an act of essentialism rooted in a view of the voice as a part of the body. While we do not advocate essentialising the body's role in determining the social meaning of voice qualities, the body is nevertheless a powerful medium through which the meaning potential of the voice is refracted." While the body is recognized as an influential mediator on the voice, Podesva also points out the relationship between physiology and ideology is often unexamined. The second stream, which I focus on for the remainder of this section, is concerned with measuring the biomechanical relationships between the different parts of the vocal system. In favouring the examination of articulatory settings from a kinematic perspective, this second stream has stepped away from examination of fine phonetic variation in speech, and questions of stance and identity completely, focussing instead on the vocal system rather than the voice itself, or the identity of the speaker beyond their body.

As highlighted by this difference in terms when referring to voice quality versus articulatory settings, the examination of vocal "defaults" occupies a somewhat liminal space between function and form, innate structure and social scaffolding. In one framing, they are the epiphenomenal result of the interaction between an anatomy with an inherent "neutral" setting, and the specifics of a given language's overall sound system and structure. In this way, they represent the most efficient, optimized, and therefore most *native-like* way to speak a given language (or dialect). However, as something that can be learned and manipulated actively by speakers, or fail to be learned even when other aspects of a second language are acquired, articulatory settings are also overlaid upon that neutral stance, altering production for some purpose (whether the purpose is efficiency or native-like production), and in that sense they are performative.

Since articulatory settings from a biomechanical perspective are seen to be a natural consequence of the interaction between anatomy and the underlying sound

[11]This should not be confused with Britto and Doyle (1990)'s (and others) use of *habitual* pitch to distinguish it from "natural" pitch, F0.

system, changing one or the other can result in changes to articulatory settings. Changes in the sound system are well-attested to alter articulatory settings across languages (Honikman, 1964; Gick et al., 2004), and dialects of the same language (Wieling and Tiede, 2017; Recasens, 2010). A series of experiments by Wilson and Gick (2014) found that, in English/French bilinguals, there is evidence of "two monolingual articulatory settings" and no evidence of a combined "bilingual AS [articulatory setting]", indicating that the articulatory settings are not just a consequence of the combined properties of speaker and sound system, but are also encoded into the language on some level. Evidence for the influence of anatomy on articulatory settings comes from differences found across age (Walsh and Smith, 2002) and individuals (Winkler, Fuchs, and Perrier, 2006; Derrick and Gick, 2021): Derrick and Gick (2021) examined how changes in speech rate prompted changes in tongue gait (categorical changes in the trajectory of tongue movement), but that the type of gait, and the speech rate which triggers a gait change, varies across individuals. They suggest that, even accounting for variation among individuals, speakers have a tendency to fall into production categories. Though much less frequent, some research has explored the possibility that the formation of these production categories and articulatory settings may be subject to social influence, or that those categories may take on extra social meaning themselves (Chela-Flores, 2006). Stuart-Smith (1999) has found evidence of different articulatory settings based on class and gender in addition to age, and Pratt (2020) has further identified articulatory settings as part of the embodied register of a "tech" persona among high school students. D'Onofrio, Pratt, and Van Hofwegen (2019, p. 212) also implicate the attachment of social meaning to articulatory settings in examining the unexpected compression of the vowel space among speakers in California.

Somewhat surprisingly, while there are numerous links drawn between vocal settings as indexing multiple aspects of identity, including gender, there is a dearth of work directly examining or attempting to describe kinematic acoustics in the framing of sex/gender, though articulation/production differences are found consistently. Even Fant (1975, p. 22) finds differences in F2 in Italian between men and women, and his explanation regarding sex/gender differences in production could be interpreted as evoking articulatory settings: "Finally, we can not quite exclude the possibility of universal "feministic" preference in vowel qualities which might have influenced the average data." (p.18).[12] While not directly examining articulatory settings by name, I present two studies that seem to explore the same phenomenon, in that they are looking at system-level influences on the perception of gender. Perception studies by Singh and Murry (1978) found that listeners use overall indications of the gender/sex of voices as the strongest cue in determining how similar two voices are from each other, and further that different features are used for judging the similarity of same sex/gender voices (such as "hoarseness" for men and segmental duration for women). Childers and Wu (1991) found that even isolated formant features carried enough overall information for listeners to correctly identify speaker gender 90% of the time, while Skuk and Schweinberger (2014), found that they were able to manipulate listener's perception of speaker gender most easily by manipulating all formants at the same time. Both studies support an influence of sex/gender on articulatory settings, in that information regarding the overall system is present even in isolated signals, and that manipulation of the overall system can shift perception. This is similar to the division proposed in the normalisation literature explored above,

[12] I note here, once more, the general framing of the male as default, in which women are described as compensating for a smaller vocal cavity (rather than men compensating for a larger one), and where the "feministic" vowel qualities are separated from the male average.

between "extraneous" speaker information as that which is present in the entire signal, and socioliguistically-"relevant" speech information as that which applies to singular variables, such as vowels.

I believe the absence of direct study of sex/gender-sourced kinematic articulatory settings in one stream, and the minimal examination of the link between the body and voice quality on the other stream, is due in part to the shared liminality of both articulatory settings and our concepts of sex/gender as existing between and across the innate properties of a body that is capable of speech, and the agency of the speaker in producing it. First, there is an assumption that "sex-specific" differences are assumed to be motivated by physiology, and therefore present in the source, the "neutral" position that articulatory settings are overlaid upon - this is the connection made by Gick et al. (2004) to explain the differences in rest position found between men and women in their study. Second, that "gender-specific" settings are seen to apply at some point afterword, where speaker agency in constructing identity, mood, or other social dimensions are overlaid on top of, but separate from, articulatory settings.

I am attempting here to collate a number of different threads of the literature regarding articulatory settings and the production/perception of gendered speech. With very few exceptions (such as Stuart-Smith (1999), Chela-Flores (2006), and Pratt (2020)), the liminality of articulatory settings as both socially malleable and arising from the inherent properties of a speaker (and language) is often ignored in favour of a focus purely on the latter, with the former being treated as a separate phenomenon. I have tried to relate this to the chapter's theme of sex/gender as occupying a similarly liminal space, and show how this has resulted in a gap in the literature. The influence of articulatory settings on listener perceptions of sex/gender can be interrogated in the same way we examine their influence on the perception of native-like speech.

2.2.4 Conclusion

This section has provided a broad examination of the differences in speech between men and women. It has examined human sexual dimorphism in the voice from a practical and theoretical perspective, in both nonlinguistic and linguistic vocalisations, from the perspective of both production and perception. In outlining the literature exploring variation as a consequence of both anatomical and social drivers, I have highlighted the ways a false division between physical and social contexts has been, both intentionally and unintentionally, imposed and legitimised over time through the forces of generalisation and abstraction. I have also examined how this false division has left a gap in the literature in exploring the liminal spaces between these sources of influence, and how our broader concepts of sex/gender beyond speech production cross similar boundaries between physiology and ideology.

It is not the purpose of this section (nor this dissertation as a whole) to provide a full and complete critique of the field of sex/gender studies within articulatory phonetics, nor offer solutions to what I present as long-standing and far-reaching issues. I have attempted in this section to provide "just enough" information about the current (and historical) landscape to be able to establish my work within it. However, it is impossible to situate my examination of nonbinary speech within the existing literature without challenging that literature's history of exclusion: contextualising the research requires problemetising the milieu, and making space sometimes requires making noise.

In this section I have also limited my scope to work from the articulatory phonetics and variationist domains. In the next section, I focus more on the concepts presented

by cultural anthropology and third-wave sociolinguistics. Like the other dichotomies discussed in this chapter, it is a somewhat arbitrary categorisation. I present this information in the way I acquired it myself: In seeking to understand concepts like performativity and embodiment from the starting point of measurable properties, the subjective, socially constructed nature of the *act* of measuring, of deciding what and who to measure, becomes apparent.

2.3 Setting the Stage: Gender/Sex and Sociocultural Linguistics

2.3.1 Introduction

The examination of sex/gender from a cultural standpoint is vast and impossible to cover exhaustively. This section is a counterpoint to the previous section, and outlines sex/gender first from the more abstract point of view taken by ancient philosophers and feminist scholars, through to the more contemporary understandings from variationist and sociocultural linguistics. Like the preceding section, is it not meant to be an exhaustive history of multiple disciplines, but to provide enough information to situate the hypotheses I present in the following chapter. Consequently, I am unable to give each area the full attention it deserves, and omit considerable detail.

I begin with a history of the terms and concepts in Section 2.3.2, the intention of which is to sketch out the shared social foundation, the concept of binary gender as it exists for my participants. Having done so, I then explore how sociolinguistics has worked within this foundation to examine sex/gender's effect on variation in three "waves", first as driven by static demographics, through constructionist approaches identifying the ideological and semiotic relationship between culture and identity as codified through language, to one facet of a complicated relationship of emergent style- and stance-making in Section 2.3.3. Finally, I provide an overview of the (small, but growing) literature specifically addressing the sociolinguistics of nonbinary identity in Section 2.4.

2.3.2 Conceptual History

I highlight here only a small aspect of the examination of sex/gender as a concept, to emphasise that the fuzzy boundaries between the physical and the social highlighted in the preceding sections are the norm historically. The way we divide or combine definitions of sex/gender to incorporate (or reject) features of biology, society, morality, performance, and others are constantly in flux. Lacquer outlines how the ancient Greeks conceived of sex as a "one-body" system (Laqueur, 1992): Aristotle referred to the difference between men and women as *contrariety* (Dea, 2016, p. 32), proposing that humans were made of different manifestations of the same "stuff", ovaries and testes were referred to with the same word, and the vagina was assumed to be an interior penis (Laqueur, 1992, p. 4). Of course, lest we assume the one-body system was conceived inherently free of patriarchy, it envisioned the masculine manifestation as superior, and viewed women as inverted, poorly-formed men, as Aristotle summarises: "We should regard the female nature as afflicted with a natural defectiveness" (De Beauvoir, 1956, p. 15). According to Laqueur (p.151), this conception shifted in the 18th century for a number of reasons, including the epistemological rise of skepticism and emphasis on the reduction of experience to observable fact (both of which served to sharpen the divide between "biological

sex and theatrical gender"), the growing political struggle for women's rights, and the cultural shift in interpreting (and policing) sexual activity as "right" or "wrong" based on reproduction: Things that resulted in reproduction were good, while things that did not were "frauds against procreation" (Foucault, 2008, p. 30).[13] Throughout the 19th century, these divisions become codified into science and law, according to Foucault:

> "[The science of sex] set itself up as the supreme authority in matters of hygienic necessity, taking up the old fears of venereal affliction and combining them with the new themes of asepsis and the great evolutionist myths with the recent institutions of public health; it claimed to ensure the physical vigor and the moral cleanliness of the social body; it promised to eliminate defective individuals, degenerate and bastardised populations in the name of a biological and historical urgency, it justified the racisms of the state, which at the time were on the horizon. It grounded them in "truth.""
>
> (Foucault (2008, pp. 54–55))

These harsh divisions are broken down and redefined by feminist scholars throughout the 20th century. De Beauvoir reconstructed binary sex/gender as asymmetrical, not representing equal opposites but identifying the masculine as the default, the central, and the positive, and femininity as the "Other", the peripheral, the negative (De Beauvoir, 1956, p. 15). De Beauvoir challenged the supremacy of biology, and famously asserted that women are made, not born, though still defining womanhood fundamentally by the biological capacity for reproduction (Tarrant, 2013, p. 171). Butler went further, proposing in Gender Trouble;

> Assuming for the moment the stability of binary sex, it does not follow that the construction of "men" will accrue exclusively to the bodies of males or that "women" will interpret only female bodies. Further, even if the sexes appear to be unproblematically binary in their morphology and constitution (which will become a question), there is no reason to assume that genders ought also to remain as two. The presumption of a binary gender system implicitly retains the belief in a mimetic relation of gender to sex whereby gender mirrors sex or is otherwise restricted by it." (Butler (1990, p. 6))

Since sex and gender are both culturally enacted in this way, and there is no action or body that exists outside of the cultural context, Butler asserts there is therefore no "true" underlying sex or gender. Gender itself is not a thing that exists, or concrete way of being, but "an identity tenuously constituted in time, instituted in an exterior space through a stylized repetition of acts." (Butler, 1990, p. 140). If the gendered body is inseparable from the acts that constitute it, there is no "doer" behind the "doing". Salih points out that Butler themself, in later editions of Gender Trouble, tempers this strong denial of a pre-discursive subject as insufficient for describing how performativity interacts with many marginalised identities (Salih, 2007, p. 60), and how in later work, they address the way queer performativities can be enacted to subvert normative expectations and "significantly deterritorialise" them (Salih, 2007, p. 63; Butler, 2011, p. 21; Butler et al., 2004, p. 218).

[13] Note that Laqueur's interpretation of history is far from uncontested. One of the most extensive of these critiques is King (2016), which primarily argues against the weight of the historical evidence for the one-body system as being the primary model for conceiving sex/gender.

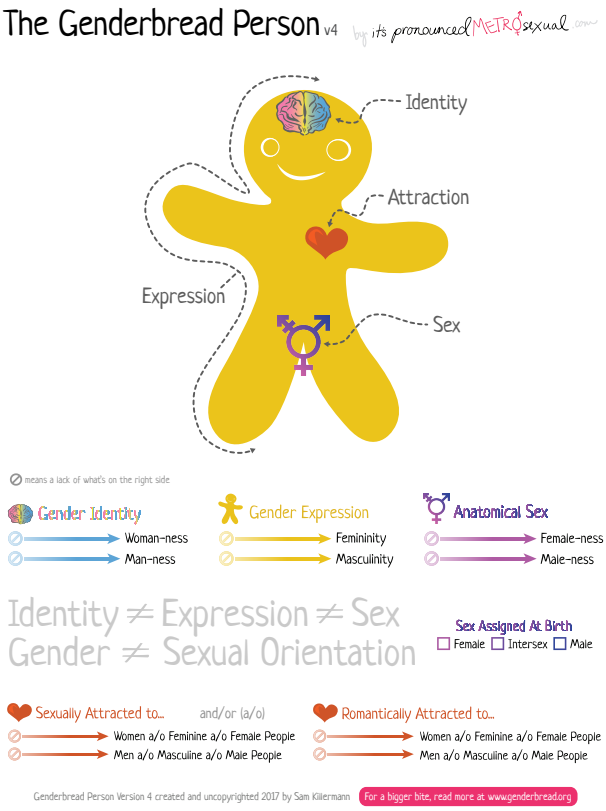


FIGURE 2.1: "The Genderbread Person" infographic intended as an introductory explanation for the different aspects of sex, gender, and attraction, that maintains a strict division between biological sex and socially constituted gender. Uncopyrighted and gifted to the commons by Sam Killermann.

Of course, the cyclical cultural dominance of one conception of sex/gender does not mean the other is excluded completely from discourse. Recent work within strongly biologically-motivated fields such as neuroimaging argues for collapsing the terms, critiquing the division as reinforcing an unproven "implicit assumption of distinctive female vs. male brains giving rise to “feminine” and “masculine” behavior, respectively." (Rippon et al., 2014). Conversely, educational materials regarding LGBT+ identities uphold the division (see Figure 2.1 for one ubiquitous example) and encourage a concept of self that isolates and defines each aspect of identity separately. Today, this adherence to a a strict divide between sex and gender, that treats sex as natural property and gender as the social consequence of that natural property, is often referred to as biological (or gender) essentialism, and its practitioners as bioessentialists, in contrast with the (social) constructionist point of view (Griffin, 2017; Chandler and Munday, 2011), which collapses them as features inextricable from the social lens that interprets them. While this dissertation takes the constructionist stance, the influence of the essentialist line of thinking is apparent throughout. This inability for society to settle on the defining features separating binary sex from gender, even as it treats the binary categorisation as an objective truth to be policed, is the context nonbinary people find themselves trying to exist within.

2.3.3 Gender/Sex in Variationist Sociocultural Linguistics

In some ways, shifts in the study of linguistic variation mirror the arguments described in the last section between essentialist and constructionist perspectives. While the changes in sociolinguistics I describe apply to many aspects of identity that are not exclusively related to the examination of sex/gender variation, the histories are deeply intertwined for the same reasons the concepts of sex/gender themselves are: The way you define categories, and the power or influence you grant one set of features or categories over another (or the direction and degree of influence attributed to the subject on category membership), is going to shape and define what it means to study the category itself.

Eckert (2012) identifies three "waves" of sociolingustic analysis, mirroring the nomenclature used in feminist literature and broadly divided by time. Levon and Mendes (2016) (and also Levon (2021)) re-articulate the three categories, maintaining some of their underlying character but re-conceptualising (and renaming) them based on their theoretical approach (Levon, 2021, p. 38).[14]

Historically, first- and second-wave sociolinguistics has tended to treat sex/gender as an underlying, immutable aspect of identity of which speech is a reflection. Along with location, age, and social class, sex/gender was a foundational category onto which variation was hung. In the correlational/first wave, pioneered by Labov (1968), these categories were taken to be relatively universal, and the interaction between socially stratified categories and language variation could be principlised, even if the exact linguistic variables changed. The constructionist/second wave challenged these principals on the grounds of their universality, focussing on ethnography, social networks, and communities of practice to examine the ways indexing between identity and speech features, or the influence and interpretation of concepts like "prestige", may vary based on local cultural context. Rather than identifying the *correlational* link between a speaker's identity and their speech, as in the first wave, more attention is paid to the semiotic processes by which linguistic difference is *ideologically* linked to identity (Irvine, Gal, and Kroskrity, 2009, p. 403). Emergent/third-wave sociolinguistics places prominence back on the dynamic, interactional, iterated nature of language and identity, and incorporates the ideological processes identified in the second wave into the creation of linguistic variation itself. Speech is conceived as one part of the wider embodiment of self that we present to the world, and that self is both presented and created, in part, through speech.

Third-wave sociolinguistics has its roots in Peirce's semiotic theory (Peirce, 1977), which describes the process of human meaning-making as a relationship between a signifier, an object, and an interpretant (Atkin, 2022). So, for example, the written word <Jacq> is a signifier, whose link to the object (the person Jacq) exists via the interpreter who reads those words. The relationship between signifiers and objects can be further categorized into three types: symbolic, iconic, and indexical. Symbolic relationships are arbitrary, iconic relationships have strong, obvious connections, and indexical relationships are somewhere in the middle, as there is some connection between the signifier and object, but the relationship is indirect. Eckert (2019, p. 754) frames this relationship more simply: "Peirce's (1960–1966) system of nested sign trichotomies has at its base three sign types, or relations between the sign vehicle and its object: symbols by pure convention, icons by resemblance, and indexes by existential relation" Changing any of the elements (signifier, signified, interpreter) changes the sign. Using the same example, if the signifier is my name, <jacq>, it is a

[14] There are other ways of describing the field which I leave aside here, for example Bell (2016, pp. 401–403)'s division based on an analysis that privileges structure, agency, or the interaction between them.

symbolic relationship. A photograph of me would be iconic, and an imprint of my shoes or picture of my glasses would be indexical. Of course, the interpreter is just as crucial to the whole as the other two elements – if you are my friend, for example, your understanding of the relationship between the signifiers and the objects is going to be different than if you only know me as the author of this dissertation and have never seen my glasses, or if you do not know me at all.

The idea of an indirect but nonarbitrary relationship between signifier and signified can be unpacked, and the conceptual distance (or indexical strength) examined. Does a sundial index the time more or less strongly than a smartwatch? Within sociolinguistics, Silverstein first described the way Pierce's concepts of icons, indexes, and symbols could be incorporated into the study of language, and specified indexical sign types as those whose "... rules of use state the relationship of mutually implied existence of a sign vehicle token[15] and certain aspects of the context of discourse." (Silverstein, 1976, p. 33). In Silverstein (2003), this is extended further to propose linguistic forms encode multiple levels of meaning-making, dubbed *nth order indexicality*. Using one of Silverstein's more accessible examples, first-order indexicality is codified in languages that make a distinction between formal and informal form of the second person pronoun (*you* in English), often called a T – V distinction, after the Latin forms *tu* and *vos*. T – V words, according to Silverstein, themselves contain a "deferential index" (p.210), indicating they encode politeness, while "enregistered honorification" is a second-order indexicality that accounts for the particular power relationship between the speaker and listener that would underlie the use of each form. Third order indexicality refers to a separate social interpretation of the variation (for example, beyond the power relationship between speaker-listener, perhaps a particular use of T – V may indicate locality, or contextual factors such as sarcasm).

Eckert's reimagining of ordered indexicality as an indexical field alters the dimensions involved. If nth-order indexicality forms a two dimensional, hierarchical relationship between form and meaning, indexical fields propose these relationships as a constellation of meanings surrounding a form, where the relationships between indexed meanings are also represented Eckert (2008, p. 454). Figure 2.2, reproduced from Eckert (2008, p. 469), gives an example of one possible indexical field surrounding a released /t/ in American English (as opposed to the standard flap). In this conception, the release of /t/ can index multiple temporary stances (in grey), more permanent qualities of the speaker (in black), and finally be linked to ideological social types or personas (in boxes).

Sex/gender, then, becomes one aspect of the indexical field, and is open to some amount of (re)negotiation by speakers and listeners. Eckert implicates the size-frequency allometry mentioned in the previous section as having an iconising effect:

> "Probably the best-known iconic pattern in English is the association of high and low acoustic frequency with smallness and largeness, respectively. [...] A variety of associations with age (i.e. children vs. adults) and gender are clearly based on size, and the acoustic frequency of both segments and fundamental frequency robustly indexes gender, and by extension sexuality, as well as class and urbanity/rurality."
>
> Eckert, 2019, p. 763.

[15] Sign vehicle token here refers to the way in which the sign is delivered to the interpretant. ie - a variable, a speech token.

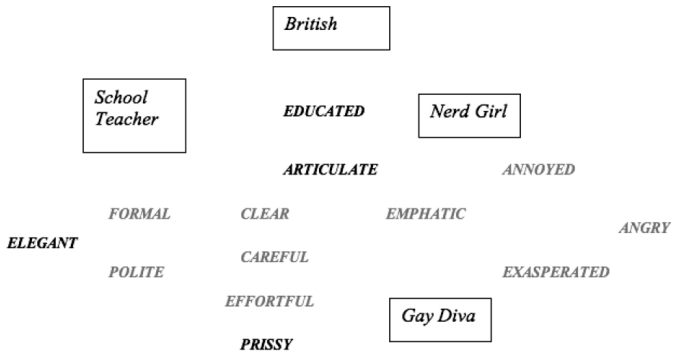


Figure 4: Indexical field of /t/ release. Boxes = social types, black = permanent qualities, gray = stances

FIGURE 2.2: Indexical Field of /t/ Release. Reproduced from Eckert (2008, p. 469) with permission. ©2022 John Wiley and Sons

The relationship between variables and their indexes, and the valence between them is constantly being updated and renegotiated by speakers and listeners. Enregisterment is the process by which variables (and the stances and qualities they index) become associated with social types (Agha, 2003).

Table 2.1 provides a summary of two ways sociolinguistic work has been categorised. There is some deviation, particularly as regards indexical fluidity, in that Eckert makes a distinction between the firmly second-wave examination of linguistic variables as a "reflection" of static identity (Eckert, 2012, p. 93), to the third-wave view in which speakers actively take part in language variation as a way to build and reinforce identity. For Levon and Mendes (2016), the emergent form of analysis arose as a result of both Eckert (2012)'s work, as well as Cameron and Kulick (2003)'s, who critique the traditional framing of "identity" itself as something, static, conscious, and consistent within an individual (Levon and Mendes, 2016, p. 5). It should also be emphasised that these approaches are not exclusive, either in time or within a single work. First wave studies are still conducted today, concepts, such as Butler's performativity (Butler, 1990) or the process of enregisterment (Agha, 2003), can straddle categories.

2.4 Setting the Stage: Nonbinary Identity in Sociolinguistics

In this section, I provide an overview of the existing scholarship within sociolinguistics examining nonbinary gender. While the totality of the body of work is relatively tiny, I still limit the scope of my presentation of representative literature here in a number of ways.

Sociolinguistics/Sociophonetics I focus here on work involving sociolinguistics, and in particular sociophonetics and the examination of phonetic features of speech (as well as the use of corpus methods). A number of recent studies have examined the linguistic structures utilised by nonbinary people, in particular the use of gender-neutral pronouns in syntax and typology (Storoshenko, 2021; Papadopoulos, 2021). As well, nonbinary speech has been examined from the perspective of nonlinguistic, sociological frameworks (such as interactionism, as in Barbee and Schrock (2019)). I consider both of these as beyond the scope of sociophonetics.

Nomenclature		Associated Concepts & Authors	
Eckert (2012), (2021)	Levon and Mendes (2016) Levon (2021)		
First Wave	Correlational	Individual Vernacular	Labov (1968), (1972)
		Sociolinguistic Interview	Tagliamonte (2006)
		Prestige	Trudgill (1972)
		Social Stratification	Trudgill (2004) and Labov (2006)
Second Wave	Constructionist	Ethnography	Bucholtz and Hall (2004), (2005)
		Social Networks	Milroy (1980)
			Milroy and Llamas (2013)
		Communities of Practice	Eckert and McConnell-Ginet (1992)
		Fractal Recursivity,	Irvine, Gal, and Kroskrity (2009)
		Iconisation, Erasure	De Beauvoir (1956)
		Indexicality	Silverstein (1976), Ochs (1992)
		Perfomativity	Butler (1990), (2004)
	Emergent	Enregisterment	Agha (2003)
Third Wave		Indexical Fluidity	Silverstein (2003)
		Styles & Stances	Coupland (1980)
			Bucholtz and Hall (2005)
		Multiplicity	Cameron and Kulick (2003)
			Eckert (2008)
		Intersectionality	Crenshaw (1989)
		Personae	Podesva (2007), (2008)
		Stylistic Bricolage	Eckert (2003)
			Zimman (2017)

TABLE 2.1: Theoretical Approaches in Sociolinguistics, Summarised from Eckert (2012), (2021), Levon and Mendes (2016) and Levon (2021). I have added De Beauvoir (1956) to the second-wave semiotic concepts of fractal recursivity, iconisation, and erasure as they both relate to ideologies sourced in identifying difference and separating a core from an "Other". Note also that, while Levon (2021, p. 38) identifies Butler (1990; 2004)'s concept of perfomativity as foundational to the constructionist approach, Eckert places their work firmly in the third-wave category, describing it as emergent in Eckert and Podesva (2021, p. 28). I have defined each of these concepts in the glossary found in Appendix A, but avoid a rote list of definitions here.

Colonial/White/Global North Englishes I focus here on the examination of nonbinary identity among English speakers in the context of the White/Colonial/-Global North hegemony.

Colonial/White/Global North Almost every paper exploring nonbinary identity has a paragraph dedicated to listing genders outside of the hegemonic, white, colonial binary: *Hijras* in Uttar Pradesh (Hall, 2013), Samoan *fa'afafine* (Schmidt, 2001), *ninauposkitzipxpe* in the oral histories of Piikáni Blackfoot (Lewis, 1941), Black Queer and QTPOC American Englishes (Burns, 2021), and a hundred others. I leave aside the vast and important body of work exploring these identities within their own context, not because I think they are unimportant, but because they are too important to be framed in this way. A cursory world-tour of citations in the scope of this dissertation presents less of an acknowledgement and celebration of the complexity of gender identity worldwide, and more of a reinforcement of the hegemonic gender binary as a cultural universal against which a diversity of identities are monolithically othered. The cultural history and nuance of these identities in context deserves more than I can give them. I reflect on this exclusion particularly in regards to the New Zealand context and takatāpui participants in Chapter 10.4.2.

English Our understanding of how language is important to the construction of nonbinary identities has not been limited to World Englishes. There is

fast-growing body of work being produced examining the construction and language of nonbinary identity in many other languages, including French (Knisely, 2020b), (2020a) and Finnish (Lindeman, 2021). In this overview, I limit my focus to work examining speakers whose primary language is English, as that is the first language of all the speakers analysed in this dissertation.

Nonbinary as a primary sex/gender There is a much larger history of work exploring gender-nonconformity and discomfort with binary gender among binary trans people[16] or more generally queer-identified participants: Calder (2020) provides an accessible and topical overview of the field from this perspective. I consider work like Hazenberg (2017)'s examination of gender-liminal spaces among trans women and Dipzinski and Abner (2021)'s exploration of the speech of queer spoken-word artists as both outside the scope of this review. I focus here on work with speakers whose nonbinary gender is asserted by the participants and considered primary by the work itself, even in cases where it may exist alongside other sex/gender identities. I also leave aside Pearce (2019)'s work on /s/ production among trans individuals, and work like it, as including nonbinary gender as a secondary classification within a larger binary paradigm, for example by extending definitions based on ASAB.

Having spent so much time dedicated to the scholarship I am leaving aside, what remains is relatively short. It is probably impossible to present an exhaustive overview of any field, even given the constraints presented, and given the pace of the emerging scholarship I'm sure it is already outdated. I provide here virtually every peer-reviewed work on the sociophonetics of nonbinary gender I was aware of at the time of writing. Dissertations and conference talks are over-represented, as might be expected given the relative youth of the field. While the remainder of this section is a list, it is a comparatively short one.

Lal Zimman. I present Zimman's work as its own section because his work, while not primarily focussed on nonbinary speakers, consistently engages with prevailing stereotypes regarding sex/gender, particularly the way masculinities are discursively constructed and embodied. I avoid listing their entire bibliography and limit myself only to those more primarily/directly addressing nonbinary gender.

Livejournal Corpus Analysis (Zimman and Hayworth, 2020; Zimman and Hayworth, 2021). This body of corpus work examines the different uses of labels and discursive identity construction among cis, trans, and nonbinary internet communities.

Phonetic Analysis (Zimman, 2021). Of particular note is his examination of the relationship between F0 and /s/ production in transmasculine speakers (both binary and nonbinary) as a process of stylistic bricolage in Zimman (2017), to which this thesis' hypotheses regarding incongruence owe a great deal.

Discourse Analysis While Zimman (2014) does not deal with nonbinary gender directly, I think it is an important inclusion within his greater body of

[16] It is probably incorrect to categorise all of the people represented by this work as "binary", and in fact it was quite common historically to deliberately collapse nonbinary subjects into binary categories (Jones, 2018), for much the same reasons given by Statistics New Zealand in Section 1.1.1.

> work because it addresses the potential for harm in arguing for the total collapse of sex/gender to a single category in regular life: "a conceptual distinction between gender and sex is crucial for the way many trans people articulate a self-identified internally-felt gender that is separate from their assigned sex. Over time, though, many trans people have become uncomfortable with a system that treats only non-trans people as biologically, naturally, male or female and implies that however a trans person might change or experience his or her body, it will never be quite like the real thing." (Zimman, 2014, pp. 20–21)

Other authors examining nonobinary sociophonetics:

Corwin (2009) and Corwin (2017): An exploration of narrative, phonetic, gestural, and stylistic practices among members of a genderqueer community in Northern California. Semiotic signals are utilised in conjunction with the surrounding environment to construct gender in different ways for different speakers based on their "durability", such that high-durability signs (such as hairline or pitch) are emphasised or hidden while low-durability signs (such as pitch contour or the use of gesture/posture) are mutable moment-to-moment.

Becker, Khan, and Zimman (2015): An examination of creaky voice that splits participant identity according to sex/hormone profile (Assigned Male at Birth (AMAB), Assigned Female at Birth (AFAB), and AFAB + testosterone) and identified gender (female/male/nonbinary). They found that sex was the only social predictor of creak.

Kirtley (2015): A PhD dissertation examining the three nonbinary transmasculine speakers in Hawai'i. One speaker identifies as a trans man, one as a lesbian, and one as māhū (a culturally-entrenched third gender in Hawai'i/Tahiti). The speech of all three changes in different contexts (such as lowering pitch during the interview with the researcher), indicating different forms of identity management and alignment/rejection of gendered (masculine/feminine/māhū) forms, while maintaining an authentic presentation of their conception of self.

Gratton (2016): An examination of Canadian nonbinary speakers' productions of (ING) based on their ASAB and the perceived safety of the environment (threat of being misgendered). Two speakers increase their use of (ING) in non-safe-spaces, indicating increased pressure to assert their identity.

Steele (2019): A Master's thesis examining F0 and /s/ among white and Black nonbinary speakers. They found that nonbinary speakers produce non-normative forms of /s/ and F0, and further that Black nonbinary speakers produce /s/ differently than white nonbinary speakers, as a reflection of a differing relationship to masculinity.

Cordoba (2020) A corpus and discourse analytic examination of 22 nonbinary speakers in the UK, exploring the "linguistic becomings" of nonbinary identity, referring to the way nonbinary speakers use language to understand and affirm their own gender via the adoption and (re)negotiation of gendered language.

Garmpi (2021) : A thematic analysis of nonbinary narratives that found participants discursively constructed their identity via semiotic processes, like iconisation, and embodied cues, like binding.

Keshav (2021): An examination of the variation in prosody guiding style-shifts in a nonbinary speaker based on changes in interlocutor, highlighting some difficulties in examining style-shifting across different speaker relationships. Recordings with the researcher present found an increased use of creak, while recordings with intimate partners found evidence of pitch accommodation.

Rechsteiner (2021): A follow-up to Gratton (2016), examining nonbinary speakers' productions of (ING) across topics. This study did not find the same shift in (ING) production based on a response to the perceived threat of being misgendered.

Overall, what emerges from this body of work is a theme of nonbinary speakers navigating complex cultural and phonetic landscapes in presenting themselves to the world. Some similarities with the (presumably) trans populations explored throughout the rest of the chapter are apparent: The importance of nonlinguistic semiotic signals such as physical appearance in presenting gender; the distress caused by miscategorisation and misgendering, and the utilisation of avoidance strategies to prevent that distress; the complicated relationship between multiple factors, such as race and sexuality, in embodying and constructing identity; the way the body influences, but does not dictate, speech.

For nonbinary speakers, however, there are other concerns and complications in presenting their gender. While there is a similar dependence on binary stereotypes to reject the incorrect binary identity imposed on them by some perceivers, there is additional pressure to construct their nonbinary identity as more than just a rejection of one (or two) binaries. Between the beginning of my PhD at the University of Canterbury in 2016, and as I write this now in 2021, the field has expanded greatly, much the same way the number of people identifying as nonbinary has grown.

Chapter 3

Defining Incongruence: Assumptions and Hypotheses

3.1 Introduction

Having presented the research questions in 1, and the necessary contextualisation via a review of the literature in 2, I now narrow the focus to forming testable hypotheses regarding the sociophonetics of nonbinary identity. I present in this chapter a proposal of one possible way nonbinary speakers may present their identity in speech, via the creation of a particular form of bricolage I call *incongruence*. In Section 3.1.1, I explain three assumptions made about the world, and why those assumptions are necessary for the creation of incongruence. In Section 3.2, I provide a definition of incongruence as a particular form of bricolage that can be utilised by speakers to manipulate speech signals in such a way that they violate binary-gender stereotypes to create the impression of something new. I also outline the conditions necessary for incongruence to arise as a possibility. In Section 3.3, I present three hypotheses, regarding the utilisation of incongruence in speech to present nonbinary identity (**H1**), the influence of the immediate context in variability (**H2**), and the tempering effect of the individual condition (**H3**) to predict differences in variation between nonbinary and binary speakers.

This strictness of this framing is to the methodologies often emphasized in the third-wave methodologies described in the previous chapter, which place emphasis on the intersectionality of speakers negotiating and creating speech- and ideologies-in-context. This is a natural expansion of the rigid macrosocial sex/gender categories proposed by the first-wave studies. Third-wave examinations of binary sex/gender began with questioning the practice of abstracting these categories and linguistic systems from linguistic behavior (Eckert and McConnell-Ginet, 1992, p. 2). In formalizing a very strict set of hypotheses to describe nonbinary speech, that cannot all hold true at the same time, my purpose is to replicate this starting point. The question of "Is there space in the classic schema to incorporate nonbinary speakers?" does not preclude criticism of those paradigms in regard to binary speakers.

3.1.1 Assumptions Inherent in the Research Questions

In examining the ways nonbinary speakers assert their nonbinary identity acoustically, I propose a series of hypotheses based on the creation of *incongruence* between signals, across signals, and in the combined presentation of multiple signals. I also predict some ways in which normative stereotypes, context, and the individual condition can shape gendered speech. These hypotheses are predicated on three assumptions about the world. The first assumption regards what I refer to throughout this thesis as the social fabric, normative stereotypes, and the shared social foundations of binary

gender, and includes both overt stereotypes, as well as what has been referred to by early sociologists as a "hidden curriculum" of gender (Jackson, 1968). This refers to the shared understanding across members of a given society, of the conceptual and concrete features of two binary genders (man/male and woman/female). The second assumption regards what I refer to as the empirical reality. It is the assumption that variation within individuals is limited by their unique experiences and physiology. The third assumption is that people, in particular the participants examined in this thesis, generally speak in a way that aligns with their (gender) identity, relative to assumptions one and two.

Assumption 1: Ideological Reality, or the Shared Social Foundation

All speakers examined within this thesis (and indeed, within the speech community of NZE speakers in Ōtautahi Christchurch, Aotearoa New Zealand) share an understanding of binary gender as an ideological reality, due to their shared social context and acquisition of patriarchal norms via explicit and hidden curricula.

While acknowledging that the definitional nuance and embodiment of sex and gender are subject to individual variation, virtually all members of this group have a shared understanding of the gender binary. All participants' experiences and upbringing contain(ed) an education in binary gender, both explicit and implied, and includes both *permission* and *policing* in the postmodern sense: Policing refers to the correction of deviance from binary norms (such as taking the "wrong" toy away), while permission refers to the way conformation to those same norms is encouraged or praised (such as via gendered compliments).[1] While, conceptually, masculinity and femininity are always open to (re)interpretation, on both an individual and macrosocial level, the ideological reality or shared social foundation is what makes that reinterpretation possible. That is, we must have a shared baseline understanding of what it ultimately means to be male or female, before we are able to change the gendered interpretation of something within that system.

This shared social foundation is wide-reaching in scope, encompassing both physical and behavioral stereotypes about what is masculine and feminine, as well as the value judgements attached to those stereotypes. For example, explicit instruction that men are physically stronger than women (that men are strong and women are weak), likewise explicit instruction that being strong is better than being weak, and instruction (either explicit, hidden, or both) that, therefore, men are stronger than women, stronger men are better than weaker men, and weaker men are more effeminate than stronger men. Of particular interest to this thesis is the assumption that acoustic features are likewise gendered, and that these features are measurable, both in general and relative terms: For example, measurements of pitch within low ranges are considered masculine, therefore a given human pitch that is lower than another is considered more masculine, all else being equal. Masculine features, such as lower pitch, are not only associated with men/males/masculinity, but masculinise the individual to some degree. Feminine features such as higher pitch, not only feminise an individual but emasculate them – to be feminine is to lose masculinity.

[1] In addition to the literature contained in the review I provided previously, also see Giroux and Purpel, 1983, Pascoe, 2011, Martin, 1998, Jackson, 1968, and Foucault, 2008 for for a deeper dive into the "hidden curriculum" of gender and sexuality. Additional exploration of hegemonic gender stereotypes in the context of Aotearoa New Zealand, in particular the ways men and women relate to the "Kiwi bloke" stereotype, can be found in Cupples, Guyatt, and Pearce, 2007 and Kuehn and Parker, 2021.

Assumption 2: Empirical Reality, or Limitations to Variation

Variation is influenced and bounded by individual experiences and physiology.

This assumption regards the ways experience shapes the variation available to individuals, both in regards to the "hard" limits of physiology, and the "soft" limits of recent experience and motivation. At its core, this assumption recognises that an individual exists inside a unique body and in a unique position within their community, and those circumstances limit the variation available to them in unique ways. For example, while a person who is 45kg may gain weight until they are 95kg, and a person who is 145kg may lose weight until they are 95kg, that does not mean they both have the same *experience* of what it means to "be 95kg." Not only does their experience of being and becoming 95kg differ, but the person who used to be 45kg may be incapable of becoming 145kg, and vice versa. Both individuals, however, have similar insight and understanding regarding the way society interprets and values people who are under- or overweight, due to their shared social foundation of body size normativity and what an individual's body size is taken to "mean" in terms of that person's strength, health, desirability, impulse control, etc. Their lived experiences differ, but lived experiences take place within the wider social context.

While I do intend here to make a distinction between "things we can do something about" (such as the people we choose to date) and "things we cannot do anything about" (such as the colour of our skin), I recognise that these boundaries are not necessarily as clear-cut as I present them, nor should "soft" limits be seen as reflection of an individual's will or desire. It is indeed possible to change many "hard" limits via surgery or dedicated sustained effort, and likewise impossible to change a "soft" limit if the pressure to act against it is insurmountable, such as the threat of social ostracising or a deeply-held spiritual belief. As with other categories discussed in this thesis, the edges are fuzzy. Challenges to this assumption are possible in regards to extreme circumstances, such as surgery, injury, or sustained dedication, I consider those to be exceptions that prove the general rule. Certainly it holds true within the time-scope of this thesis and these particular research participants. The "soft" limits to variation as a topic came up multiple times over the participant interviews, and is explored in detail in the sections where it is most relevant to each participant. This assumption recognises the effect of the vastly different histories over which nonbinary identity is iterated for each participant.

Assumption 3: Speech Encodes (Gender) Identity

People generally speak in a way that aligns with their gender identity.

This assumption is self-explanatory, and is the basis of identity presentation in speech. However, it is worth highlighting two particulars as they relate to this thesis. The first is that the participants in this research have a desire to present themselves as, and be seen as, the gender they were recruited as (nonbinary, male, or female). This can be inferred from their responding to recruitment materials and agreeing to participate, which aligns with assumptions made by virtually all linguistic research involving participants divided by gender. For nonbinary speakers in particular, I go into more detail in Chapter 4.3.1 regarding some of the ways they exhibit a strong intention to present themselves as nonbinary. I am confident that my participant pool excludes both "nonbinary speakers with no desire to be acknowledged as nonbinary" and "binary speakers who desire to be acknowledged as nonbinary." I make no assumptions, however, as to whether those populations actually exist in the wider world, nor discuss them further here.

The second particular of this assumption is that "gender is encoded in speech" holds true regardless of whether or not there is consistency across participants regarding what, specifically, constitutes "speaking in a way that aligns with one's gender identity." That is, a person does not need to know *how* their speech aligns with their gender identity, nor even the relationship between gender identity and speech, to speak in a way that aligns with it. The hypotheses in the following section address how this can be done. I have deliberately avoided using language here that implies active agency.[2] For the purposes of this thesis, I consider the agency involved in self-identifying one's gender to be the driving force behind the speech produced, unless otherwise expressed, rather than a conscious decision regarding speech at all.

3.2 Proposal: Incongruence

3.2.1 Introduction

Given the assumptions stated above, I hypothesise that nonbinary speakers can present their identity via incongruence. The following sections outline a formal definition of incongruence as it is used in this thesis. It also presents a "recipe" for incongruence, outlining the elements that must be present for its creation. Finally, the hypotheses are formalised, and I present the predicted forms that incongruence can take in different situations, regarding individual acoustic variables produced by single speakers, variables across speakers, and bundles of variables within individual speakers.

3.2.2 Definition

INCONGRUENCE : A mismatch between a combination of variables associated with some identity and the actual experience of those variables, arising from the identity itself being unrepresented in the association.

When the overall impression of signal (be it multiple instances of a single variable or across multiple variables) is at odds with the expected impression generally associated with that signal in regards to known identities.

Incongruence can arise between multiple instances of a single variable, or as a property of multiple variables interpreted in tandem.

This thesis is concerned with *gender incongruence* in speech, a type of incongruence that arises when the speech produced by a single nonbinary speaker cannot be reliably interpreted as belonging exclusively to one of the two binary gender categories (male or female) allowed within the normative social context[3]. Incongruence, and even incongruence in speech, is not limited only to the acoustic realm, but involves the overall perception of an entire individual. The complexities of the way we perceive identity holistically and as a multi-modal process is beyond the scope of this thesis. However, I discuss one example of the influence of physical presentation in Chapter 10, where I examine the effect of one participant's use of makeup on their speech.

[2] Though some participants *do* make explicit claims of agency themselves, regarding their motivation and desire when speaking. I have reported these whenever possible/relevant throughout this thesis.

[3] While the indexical relationship between linguistic forms and sex/gender is technically non-exclusive, as per Ochs (1992, p. 341), in that these forms are never exclusively used in practice by only one group or the other, I refer here to the "core social meaning" that links forms to either men or women.

I take the word incongruence from its definition within chemistry, where it refers to a process in which a substance undergoing a state change (such as melting or vaporisation) results in a change in composition. For example, a composition of forsterite and quartz, when heated under specific conditions, undergoes *incongruent melting*, and upon cooling/crystallisation produces pure fosterite and pure enstatite with no solid quartz remaining (Nelson, 2011).

$$Mg_2SiO_4 + SiO_2 = 2MgSiO_3$$

$$Fo + liq = 2En$$

The base elements involved in incongruent phase transitions do not change, but their composition is rearranged by a particular application of external forces, such as heat, resulting in a different substance. This preservation (albeit recombination) of the starting elements in the end product is the feature I wish to highlight. The quartz may disappear, but the elements of quarts, its constituent parts, still contribute to, and exist in, the product. I hypothesise that nonbinary speakers, working within the constraints of possible speech productions and normative ideologies about binary speech productions, are able to "rearrange" those acoustic elements to create something new perceptually, different from the original composition, but still a natural result. Drawing from the sociolingusitic literature, incongruence is a specification and extension of the concept of stylistic bricolage, as introduced by (Eckert, 2003) (and see Zimman (2017, p. 46) for more specific examinations of bricolage as it is utilised by speakers within the transmasculine community).

Within linguistics, *incongruity* is often used across disciplines as a non-specific term to refer to a mismatch between expectations and outcome: For example, a sentence with an incoherent subject (Valderrama et al., 2021, p. 97); the use of minority languages on signage at global fast food franchises (Hornsby, 2008); and collocations that differ across two or more languages (Hashemi and Eskandari, 2017) have all been described as incongruent in the literature. In syntax specifically, *congruence* as a technical term is often used interchangeably with *concord*, and refers to a grammatical agreement between elements in a sentence (such as a subject and its predicate), and therefore incongruence in that context refers to some ungrammatical, disharmonious breakdown between the involved elements. While the term incongruence, as I use it, is obviously related to the broad concept of assuming a set of shared "norms" that are open to violation, syntactic incongruence implies something has gone *wrong*, whereas chemically-incongruent melting (and the way I intend incongruence to be interpreted conceptually in this dissertation) does not.

3.2.3 A Recipe for Incongruence

For incongruence to be an option in identity construction, the following conditions are required:

1. Shared social context across interlocuters regarding the expected cues to known identities. (In this case, binary gender and the multiple cues involved in creating congruent male or congruent female presentation, a first-order indexicality between sex/gender and speech.).

2. Speakers whose identities do not fall inside of those known/recognised in the shared social context, as well as an understanding of the ways in which the new identity is related to the known. (In this case, nonbinary people and the

understanding that it is the social categories of "male" and "female" that they are not represented by. This may encompass speakers with a second-order index that isn't shared by listeners, or *n+1* indexicalities that have not been enregistered yet.)

3. A desire on the part of a given speaker with that identity to have it be known and present it.[4]

4. The ability on the part of speakers to manipulate the signal.

When all of these conditions are present, incongruence can be created via signal manipulation, subverting the expected norms in a process of bricolage to (attempt to) create the perception of something new. Incongruence exists as a specific subset of bricolage precisely because it requires these four specific elements to arise. Each of these conditions is necessary for incongruence. If there is no shared social context (1), then signals cannot be meaningfully attached to known identities in a way that makes carving space for *unknown* identities possible. If (2) does not exist, then there is nothing to represent (in the case of all speakers falling inside the known categories), or there is no way to reference known categories in relation to new categories (in the case of nonbinary identity being recognised as a sex/gender identity). Without a desire to be recognised (3), or the ability to manipulate the signal (4), there is no impetus or ability to create incongruence in the ways hypothesised.

The conditions that may drive speakers to produce incongruent signals are not limited to gender. See, for example, Stuart-Smith et al. (2011, p. 20) where a unique "Glaswasian" identity seems to be driving realizations of /l/ that are clear relative to other Glaswegian speakers, but still darker than other English accents. Kirkham (2015)'s study of the "indexical dynamics" (p.648) of ethnicity, social class, and local orientation among schoolchildren in Sheffield likewise provides evidence of the conditions required to drive sociolinguistic change towards the creation of a new identity formed from the incongruent combination of ideologically opposed parts (such as insider/outsider and working-/middle-class). The methods by which incongruence is created, and the ways signals may be perceived as (in)congruent, are unique to the relationship between the new identity and the identity (or identities) already codified in the shared social context.

I step away now from the general conditions required to make incongruence possible, and turn instead to the specific way incongruence can manifest as regards to sex/gender. A nonbinary speaker may present their nonbinary identity in speech via the manipulation of (acoustic) signals to create either *ambiguity* or *dissonance* with reference to listener expectations of (binary) sex/gender. It should be noted explicitly here that *only* nonbinary speakers are expected to utilise gender incongruence in the manner proposed. While binary sex/gender in all its intersectional complexity may not be represented equally and well in the shared social context for all binary speakers, it is still accounted for in a way that nonbinary identity is not, that excludes it from meeting the conditions for allowing incongruence.

Listeners[5] have gendered expectations in regards to speech (and speakers) they reference. The expectation is that most/all signals they receive from a single speaker,

[4]While I think it may be possible, and even expected, for some kinds of "mixed signals" to arise in ways that are not part of identity management, I would not consider them *incongruence* as such.

[5]I mean here not only *most* listeners, but also *all* listeners in regards to the general expectations based on experiences and the shared social context. Obviously nonbinary speakers, as well as many speakers with experience with people with nonbinary genders, may have slightly different or expanded expectations, but they will still have an understanding of the normative gender binary.

from multiple variables and modalities, will be interpreted as either primarily masculine-coded or primarily feminine-coded, and those that are cross-coded or ambiguous are rare outliers. Therefore, the overall presentation of a given speaker is going to reliably be one of two binary categories (man or woman).

If the signal is incongruent, the perception on the part of the listener is that the overall presentation is neither reliably male, nor reliably female. This can manifest in two different ways.

Ambiguity: The individual variables, signals, and overall presentation are *neither* reliably male nor reliably female. The overall impression is one of uncertainty.

Dissonance: The signal produced by a single individual is reliably male *and* reliably female. This may apply within individual variables (one variable contains both prototypical male and prototypical female tokens) or across variables (one variable is reliably prototypically female while another variable is reliably prototypically male). The overall impression is surprise or confusion.

There are further implications about the different ways ambiguous or dissonant incongruence can affect perceivers, the differential weighting of single cues, and as manifestations of speaker intention. For the most part I avoid addressing these implications except in those cases where participants have explicitly addressed them themselves. However, I think it is an interesting question to explore experimentally in the future.

There is another underlying assumption here, that percievers categorise human gender as being singular and holistic to a single human. This is non-controversial and aligns with a number of assumptions about the way we categorise whole objects (such as humans) and assign exclusive labels (such as "boy"), at least in acquisition (American Psychological Association Dictionary, 2022; 2022). However, it is also obvious that, post-acquisition, humans are capable of, and often do, compartmentalise sex/gender into different categories or domains when applied to a single individual: For example, descriptions like "male [soul] in a female body", "a woman's voice but a man's propensity for swearing", "hyperfeminine personality but built like a [male] linebacker" are all easily parseable by listeners. I think these forms of cross-gender descriptions work in part *because* they violate a kind of "one body = one sex/gender" assumption.

3.3 Hypotheses

3.3.1 Introduction

I hypothesise that nonbinary speakers will create acoustic incongruence to express their nonbinary identity in speech (**H1**). The utilisation of incongruence is tempered by two general forces that predict differences in production between binary and nonbinary speakers. These are the immediate context (**H2**) and the individual condition (**H3**).

The three hypotheses stem broadly from the three sources of contextual influence proposed in the last section. The shared social foundation presents the baseline stereotypes upon which incongruence can manifest acoustically, addressed in **H1**, as well as predicting differences in variability by context as predicted in **H2**. The constraints of the individual conditions leave room for detailed qualitative analysis of variation found in individual speakers, as proposed in **H3**.

3.3.2 Hypothesis 1: Acoustic Incongruence

Nonbinary speakers can create incongruence in speech either within variables or across them, via either *ambiguity* or *dissonance* with binary gendered expectations. Incongruence via ambiguity involves the use of signals that, when examined in isolation or together, cannot be reliably categorised as feminine or masculine. Incongruence via dissonance is created when a single source (speaker) produces signals that, in isolation, conform to binary masculine *and* binary feminine norms. I make a distinction in this section between categorical and gradient signals. A signal is categorical if the set of possible productions for it is discrete and countable, while gradient signals are those to which each individual member is located along one or more variable ranges/values (Myers, 1995). As we shall see later, the distinction between categorical and gradient signals, much like the distinction across gender itself, is highly dependent on context, including the different weightings granted to perception vs. production, expectation vs. novelty, and intent vs. impact. However, the utility in this approach is that it makes predictions of each variable easier to explain, and highlights the relationship between individual variables and whole-system incongruence.

Within variables where difference falls along a gender binary, incongruence is expected to manifest differently depending on whether the variable is categorical or gradient. In categorical signals, incongruence via dissonance is created via equal frequencies of both categories, while incongruence via ambiguity is created via the utilisation of a new variant. In gradient signals, incongruence via dissonance is created via a bimodal distribution of tokens spanning both masculine and feminine distributions, while incongruence via ambiguity is created with a single distribution of tokens centralised between masculine and feminine norms.

Incongruence can also be created across signals via dissonance, where a speaker's individual variables each align with binary gendered norms, but whereby some variables pattern with male norms and some with female norms. It is not possible to create incongruence across signals via ambiguity separately from incongruence via ambiguity within single signals. It should be noted explicitly that the hypotheses address different strategies, and it is impossible for all of them to hold true at all times. A single speaker cannot produce incongruence via ambiguity *and* dissonance within a single signal, and a speaker creating incongruence across signals cannot also create it within individual signals. Ambiguity and dissonancee are two different paths to reach the same destination.

Hypothesis H1a – Categorical Signal Incongruence – *Given a single categorical variable, nonbinary speakers can communicate identity via **dissonance**, by producing tokens at rates between those found for binary male and binary female speakers, or via **ambiguity**, by producing a variant not found in either men's or women's productions.*

Figure 3.1 provides a visualisation of how this hypothesis would present, using an idealised dataset. Incongruence via dissonance is achieved in categorical signals via the production of both tokens perceived as masculine and tokens perceived as feminine, at roughly equivalent rates. Incongruence via ambiguity may be achieved via the creation of a "new" variant, that is not reliably associated with either men or women, but acoustically plausible during speech.

Hypothesis H1b – Gradient Signal Incongruence – *Given a single gradient variable, nonbinary speakers can communicate identity by producing incongruence either via **ambiguity** in the form of more centrally-distributed tokens compared to binary male and binary female speakers, or via **dissonance** by producing tokens in a bimodal fashion, where a single speaker's productions are within both binary male and binary female ranges.*

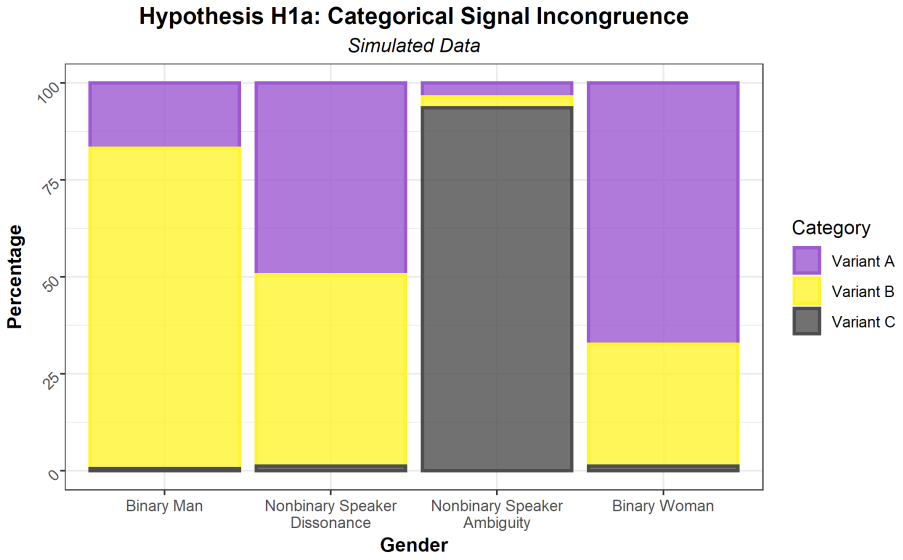


FIGURE 3.1: **H1a**: Categorical Signal Incongruence: Simulated distribution of token counts between variants of a given categorical variable.

Figures 3.2 and 3.3 present **H1b** using idealised data. An ambiguous distribution of a gradient cue more abruptly challenges the ideological reality of binary gender: Most tokens are distributed away from both typically- and hyper-masculine and typically- and hyper-feminine productions. An ambiguous distribution can be argued to be the most efficient expression of difference for the least amount of articulatory effort (one of the proposed drive behind articulatory settings in kinematics). Finally, this distribution can be seen to more closely resemble observed differences between binary men and women, which are relatively normal distributions shifted across some measurement space.

However, incongruence via dissonance may also be motivated theoretically. Under a model of speech perception and production based on direct storage, like exemplar theory (Johnson, 1997), speakers' experiences of speech are made up of two unimodal distributions that exist at different parts of a shared perceptual space. In selecting tokens from both distributions for production, a nonbinary speaker creating a dissonant distribution more closely represents the speech exemplars actually experienced by a given listener (Todd, 2019), whether you consider exemplars as encompassing direct experience only or as also representing popular discourses and sex/gender stereotypes (as well as the influence of the latter on the former). This also aligns with an analysis that accounts for the role of the social foundation in categorising and weighting exemplars based on group membership (Coats and Smith, 1999; Smith and Zarate, 1992), in which something that is both "not male" and "not female" at the same time does not exist in one's perceptual experiences – one can imagine the gap between masculine and feminine distributions across some measurement as a "perceptual blind spot" similar to the boundaries in categorical perception. While articulatory settings are presumed to favour a conservation of effort, the findings from an examination of articulatory settings on code-switching bilinguals (Gick et al., 2004) suggest a dissonant distribution may represent a rapid change between "masculine" and "feminine" settings. Finally, in relation to a speaker's individual condition, the two peaks of a dissonant distribution may represent what can be conceived of as a "default" mode, more closely resembling that speaker's production space before embracing a nonbinary identity, and an "effort" mode, in which presumed listener

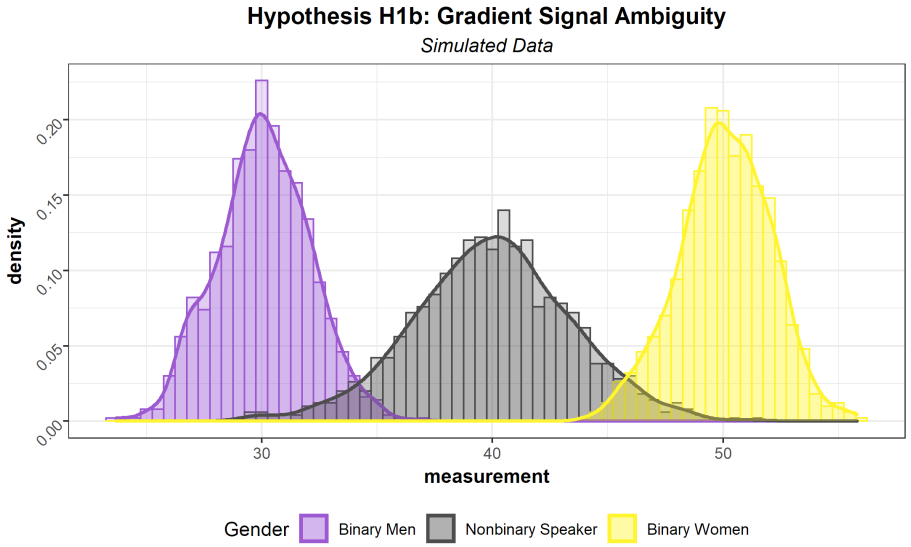


FIGURE 3.2: **H1b**: Gradient Signal Ambiguity: Simulated distribution of variation of a given gradient signal, with a nonbinary speaker producing incongruence via ambiguity.

perceptions are directly targeted for violation, rather than the sustained effort towards a singular target distribution, as found in the ambiguous case.

Hypothesis H1c – Multiple Signal Dissonance – *Given multiple signals, nonbinary speakers can communicate identity by producing incongruence via dissonance* ***across*** *signals.*

While a single variable examined in isolation may pattern explicitly with one binary gender or the other, incongruence can arise from the relationship between multiple variables: Multiple "binary gender-congruent" signals from a given nonbinary speaker examined together will show a dissonance between those signals. This is illustrated using idealised data in Figure 3.4. While it is expected for binary speakers that all signals taken together will exhibit a strong trend towards their binary gender, this is unexpected behaviour for nonbinary speakers. Multiple signal ambiguity is not impossible, but it is epiphenomenal, as it arises from multiple instances of gradient signal ambiguity.

This hypothesis can only be examined from a holistic approach in which a given speaker's variables are all examined together.

3.3.3 Hypothesis 2: External Context

Hypothesis 2 – *Immediate context has a greater influence on nonbinary speakers than binary speakers.*

It is expected that nonbinary speakers will exhibit more variation across contexts than binary speakers. This is proposed to be caused by the relative recency of "nonbinary" as an available label for participants – more recent, in fact, than their relationship to many of the people they are speaking to. The degree to which gender incongruence in speech arises is hypothesised to vary more based on topic and interlocutor, compared to the variation found in binary speakers.

3.3.4 Hypothesis 3: Individual Condition

Hypothesis 3 – *Variation is bound by the individual circumstances of each participant. Individuals with shared socially-encoded circumstances, such as age and binary gender, will*

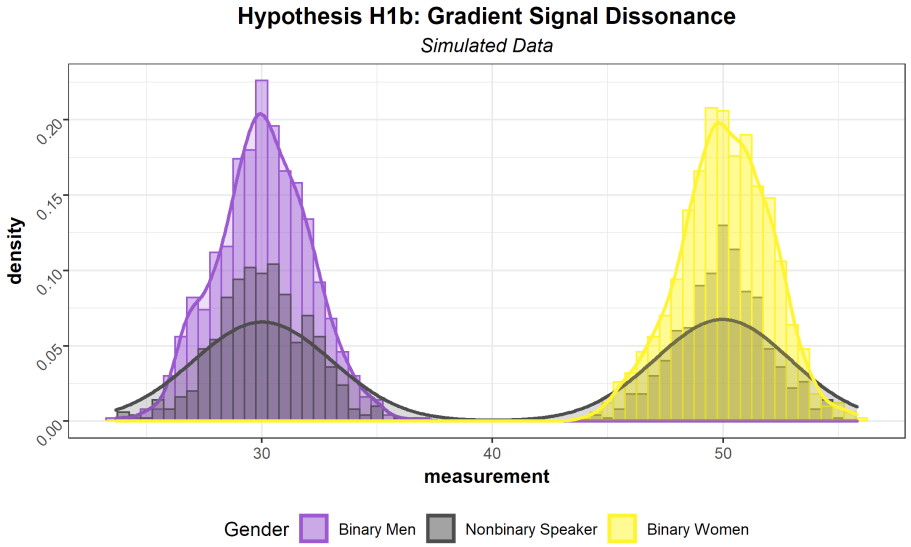


FIGURE 3.3: **H1b**: Gradient Signal Dissonance: Simulated distribution of variation of a given gradient signal, bimodal distribution

speak more like each other.

For nonbinary speakers in particular, the types of incongruence utilised will be bound by their individual condition (most obviously physiology, but also by their social history). This is true of binary speakers as well, but is expected to be more similar between same-binary-gender participants than between nonbinary ones. As discussed in the previous chapter, this is a simplification of the complex and myriad ways binary speakers can "do" gender. This hypothesis proposes that binary speakers will sound more like their same-binary-gender peers, compared to other gender groups.

This hypothesis makes two explicit predictions:

1. Due to greater similarity in the individual condition, less variation will be found within binary gender groups than within the nonbinary group.

2. Variation within the nonbinary gender group will conform to each speaker's individual condition, in terms of identity, history, and intention.

The first of these predictions is analyzed quantitatively for each variable analyzed via the question "Do nonbinary speakers vary more in production than binary speakers?". The second is addressed via the use of *Spotlights* within each analysis chapter, which examine individual speakers or small groups of speakers qualitatively.

3.4 Conclusion

In this chapter, I outlined three basic assumptions I make about the world to formally address the research questions. I then defined *incongruence* as a technical term for a form of stylistic bricolage I predict that nonbinary speakers utilise in asserting their identity as one outside of binary gender norms. I proposed two ways incongruence may be created, ambiguity and dissonance, and outlined three hypotheses regarding the creation of incongruence via speech, and the mitigating factors of external context and individual condition that shape it. In the next chapter, I outline my

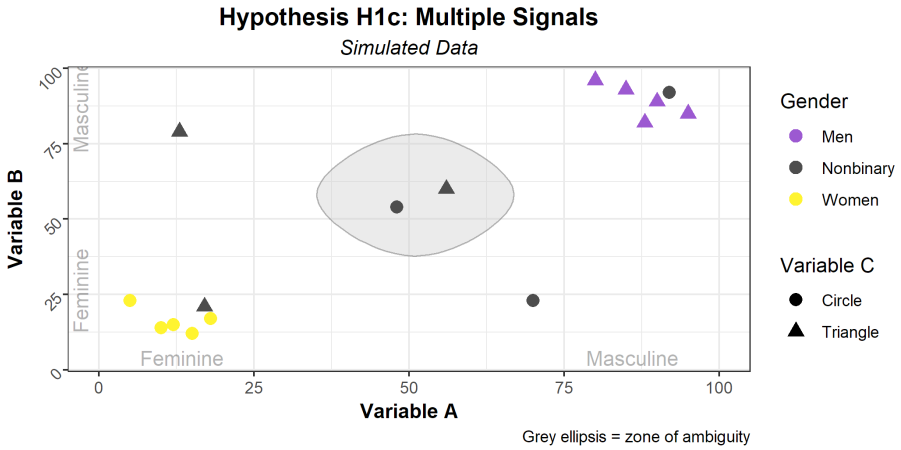


FIGURE 3.4: **H1c**: Multiple Signal Dissonance: Simulated distribution of variation multiple signals. Variables A and B, along the X- and Y-axis, represent the values of gradient variables, such as pitch or speech rate. Variable C, illustrated using shapes, represents a categorical variable, such as the presence/absence of voicing or aspiraton.

research methodology for studying incongruence, including some external context and individual factors that shaped the creation of the RAINBO Corpus.

Chapter 4

Methodology

4.1 Introduction

This chapter outlines the research methods involved in recruiting participants and collecting the data that form the Recorded Audiovisual Interviews with Nonbinary and Binary Orators (aKa RAINBO) Corpus. I begin with an explanation of my motivations in 4.2, and outline the greater context in which the recruitment took place in 4.3. In Section 4.4 I provide the data collection procedures. Section 4.5 describes the process of turning the raw data into a usable corpus. I conclude in 4.6.

4.2 Motivation

Creating a corpus, particularly one with the depth and scope of RAINBO, is not strictly necessary in examining the speech of a population. I was motivated to collect and organise data in this way primarily due to the emergent status of nonbinary identity. I approached data collection from the point of view that, some years from now, there may be a "nonbinary voice" with identifiable speech features, similar to the way sibilant hyperarticulation became a marker of gay male speech (Munson and Babel, 2007; Hazenberg, 2017). I use the data in RAINBO to address the hypotheses and research questions, but the methodology I describe in creating the corpus is much wider in scope, as it is intended to be a "living" corpus, that is accessible long-term and is open to new additions.

To address this goal, it was important for me that RAINBO was (presented in alphabetical order):

Anonymous This was important for ethical and safety considerations, both, but since the intention was to grant relatively wide access and to allow new speakers to be added, as well as allow speakers to be comfortable in talking about their attitudes regarding something quite contentious and politicised at the time, it was important that steps were taken to anonymise participant data.

Expandable / Perpetual Being open to new members over time allows the possibility of a diachronic examination of nonbinary dialect as it develops in Aotearoa New Zealand. While lofty, the hope that RAINBO might one day be an "origin point" for nonbinary speech in Aotearoa New Zealand was a driving motivation to collecting as much information as possible in a way that is intended to remain accessible to future work and additions.

Inclusive I chose to include binary speakers in RAINBO for two reasons. The first was very simply as a control population, so there is something to track the emerging nonbinary dialect against, for the timepoint the data was collected. The second,

and more abstract, was to serve as an examination of the way attitudes (as well as speech, and visual presentation) may change over time as well.

Multimodal While my primary intention was to examine acoustic data for this dissertation, I felt it was important to include a visual component in the corpus, due to the multimodal nature of gender perception.

Some of the information in RAINBO collected to meet these goals, described in this chapter, are not used in the analyses presented in the following chapters. This includes the selfie data, some speaker information such as Socioeconomic Status (SES), and higher-level organisational classifications (such as topic-tagging and timestamping). This was partially due to time constraints and participant homogeneity, as I speak about in Chapter 10. However, the intention was always to cast a wider net and collect more data than strictly necessary, to facilitate future research.

4.3 Research Context

4.3.1 Recruitment

A deeper exploration of the research context is necessary here, as it heavily influenced the sample of participants recruited and the tone of the subsequent interviews. Data collection began in March 2017, and ran for approximately two years. In May of 2016, the administration of then-president of the United States of America, Barack Obama, issued a directive to schools, asserting that transgender students be allowed to use the toilets whose signage aligned with their affirmed sex/gender (Davis and Apuzzo, 2016; Emma, 2016). This directive launched an issue with an already contentious history (Cavanagh, 2010; Herman, 2013) into the centre of the global media cycle. Multiple news articles and editorials appeared in newspapers in Aotearoa New Zealand (Holland, 2016; McSweeny, 2016), Australia (Koziol, 2017), Canada (Freeman, 2016; Lamoureux, 2016), the UK (Yorke, 2017), and across the rest of the English-speaking world. A year later, in February 2017, the newly-elected president Donald Trump rescinded that directive from the previous administration, re-igniting the debate in the public sphere (Peters, Becker, and Davis, 2017). A few months later, in August 2017, a bill was proposed to the New Zealand Parliament to allow sex/gender markers on documents to be changed via self-determination (Births, Deaths, Marriages, and Relationships Registration Bill 2017 (296-2)), localising the discussion and creating an extended discourse specifically in NZ that continues through to the time of this writing.[1] The political and media climate surrounding issues concerning Lesbian, Gay, Bi, Trans, and Others (LGBT+) in general, and "bathroom bills" in particular, were unavoidable, and these topics were immediately accessible to virtually everyone in Aotearoa New Zealand, including the participants of my study, most of which are university students.

Physical recruitment materials for nonbinary participants were regularly vandalised or had the contact information removed, and digital recruitment notices became platforms of political and social discourse. The University of Canterbury's student association runs a Facebook page that is widely used by current students, alumni, staff, and other members of the community. The scope of this group is equally wide, covering everything from flat hunting posts, questions from prospective students, community watch notices, and general discussions. It is also the *de facto*

[1] This bill passed its second reading four years later, in August 2021, mostly due to extended periods of deferral for collating feedback from the public and special interest groups (Smith, 2021).

method of electronic recruitment for research conducted by students at the University of Canterbury. Unlike paper recruitment notices, which can be easily rendered useless, electronic posts can only be removed by a limited number of people (Facebook itself, the original poster, and the group moderators/administrators). Facebook posts can also be interacted with in myriad other ways, both via direct comments, as well as indirect reactions in the form of "likes"[2] and shares. Facebook's system facilitates and encourages interaction as a design feature, and the more "engagement" a post has in the form of comments, likes, shares, and even how long a person looks at it, the more likely it is to appear in front of other users (Petre, Duffy, and Hund, 2019). The atmosphere on the student union's Facebook group reflects this technical rewarding of "banter", taking on a tone that is casual, often "edgy", and mostly unmoderated – for the most part users are able to post what they like without repercussion, and often will comment on posts with jokes, commentary, or tagging friends to draw their attention to content. This last feature is a double-edged sword for researchers – those who do not meet the recruitment criteria are able to tag friends who do, but are also able to tag their friends who *don't*, as a way to imply they do as a joke.[3]

FIGURE 4.1: A screenshot of a Facebook recruitment notice for the study, and the first response invalidating nonbinary gender.

Recruitment notices I put on facebook for nonbinary participants inevitably attracted multiple responses and reactions. One example is shown in Figure 4.1, which is an anonymised screenshot of one of my early recruitment posts, taken about an hour after it was posted. The post itself has four "likes" (thumbs up reactions), mostly from my own friend and contacts, who were notified I posted in the group, and the first response invalidating nonbinary genders has 22 "likes". Over the next few hours, the subthread under that response received dozens more posts of support, refutation, opinion, debate, and memes, replaying the wider conversation that had been occurring across media platforms for the last year. Throughout the engagement, though the number of responses was evenly split (or slightly in favour of pro-LGBT+ messages) the silent majority represented by "likes" was overwhelmingly on the side of dismissive or antagonising posts.

Aside from this initial screenshot, very little proof remains of this engagement, as many of my recruitment posts (including the one in 4.1) were deleted outright

[2]Recruitment predated Facebook's implementation of other reactions/emoji, such as hearts and laughing faces. Thankfully.

[3]As of June 2019, stricter moderation practices have been implemented within the group, due to the same problematic "joke tagging" I have described being rampant on "community watch" type posts about suspicious behavior, robberies, drugged drinks, etc.

or heavily moderated by the University of Canterbury's Student's Association to remove transphobic/homophobic comments and jokes. These circumstances, both the initial negative and invalidating response, and the heavy-handed moderation that followed, directly affected the data collection in the following ways:

1. Potential participants to the study saw their gender (and broader identity) immediately invalidated, and the current media debates about their right to exist localised further by becoming a point of public discourse (and, eventually, censorship/silencing) between their peers.

2. I received a number of messages from individuals with no intention to participate in the research. These ranged from relatively benign transphobic memes to time-wasting questions to, in one case, a picture of genitalia.[4] These frivolous messages incurred a small cost in terms of time and energy.

3. Having a post deleted and reposted is more likely to trigger Facebook's spam protection and reduce its overall visibility, limiting the number of potential participants that actually see it. Having posts deleted also made it difficult/impossible to follow-up on harassment with official university channels, as links to distressing content were deleted before being seen by relevant authorities. It is also impossible to track which of the nonbinary participants in the study saw which posts (including the recruitment notice itself and accompanying commentary).

4. The deletion of some posts and comments but not others (and the subsequent questioning of the deletions) caused confusion and led to some genuine questions and recruitment tags to be disjointed across posts, and further fractured some the topics into more general debates regarding LGBT+ rights and "censorship vs free speech".

These conditions had an effect primarily on the final participants recruited and, secondarily, on my research methodology as a whole. In the first case, most of those who agreed to participate did so despite the potentially-distressing and very public judgement and conversation immediately presented by my recruitment materials. It is easy to imagine a person who may have considered participating backing out after seeing the behavior of their classmates and peers, or being made uncomfortable by being tagged to a post where others were tagged in as a joke. Nonbinary participants that agreed to participate did so despite this wider unfriendly conversation surrounding the recruitment. While I would never frame this recruitment context as a positive thing and I wish it did not happen, I can be relatively confident that those who *did* choose to participate in the research share something significant – a core sense of their nonbinary self, a desire to be seen/heard, or some personality trait or attitude that extends beyond having a nonbinary gender. Ultimately, because of this, it is reasonable to assume that the nonbinary participants in this study have a desire to be seen and heard as nonbinary, which may manifest to a greater degree than those with less desire to put themselves through the hassle of the recruitment process.[5] From an organisational standpoint, this unfriendly recruitment context compounded an already-difficult task. I originally planned to find 12 nonbinary participants (divided

[4] While I do not want to exaggerate the overall number of messages, I also do not want to minimise their impact. Even one dick pic is too many.

[5] Though, as I discuss in 10, this may have been a bit of a double-edged sword in that the participants most comfortable in their nonbinary identity may be the least likely to exhibit my predicted hallmarks of "emergent" identity in speech, such as sensitivity to contextual variation.

roughly in half between those whose gender identity was static and those whose gender identity was mutable) in addition to 12 binary men and 12 binary women. While this was perhaps an overly ambitious number from the outset, given the time involved in recording/transcribing each participant and the numbers involved,[6] I believe the difficulty was exacerbated by the way facebook buries reposted materials. One participant explicitly mentioned that they were tagged to a post at a time coinciding with the first recruitment notices, but the link had disappeared and they could not find it again until I had posted almost a year later.

It is impossible to ignore, but also impossible to quantify, the effect this response to my materials also had on my own engagement with participants. While there are undeniable benefits to researching nonbinary language while being nonbinary myself, and I can more clearly empathise with how the recruitment process may have looked and felt to potential participants, it also means that the targeted criticism is more personal. The handful of false messages also led to a shift in tone and increased caution with potential participants – I did not disclose my gender or other personal or contact information, for example, where I might have done so in other research. I erred on the side of caution if someone's contact information felt "off", or if they asked for information like my home phone number. Due to this, it is possible that I turned away legitimate participants due to this increased suspicion, becoming a gatekeeper of identity myself.

This context also altered the way that I approached binary participants. Unsurprisingly, the recruitment notices seeking "men and women" attracted very little engagement outside of potential participants. Being relatively new to Aotearoa at the time and enjoying an incredible amount of support from my department, I was somewhat naive and assumed the attitude of the undergraduate population would be similarly supportive. Due to the deletion/moderation of the posts, I was also nervous because I had no way to know if the binary participants were the same people responsible for some of the "jokes" on the nonbinary notices. A handful of responses do not represent the baseline attitude at the University of Canterbury or Aotearoa New Zealand in general, but I was genuinely shocked at the responses to my first recruitment post, and it shifted my "baseline" assumptions about the attitudes of the non-LGBT+ student body. I am sure this had an effect in the way I engaged with the binary participants during the course of the research, even if it is impossible to know exactly how.

Environmental Context

The section above described some of the difficulties in public and semi-public recruitment of non-binary participants. The physical environment of the research itself offered additional challenges. Throughout the winter months in 2017 (June-September), the building the interviews were conducted in underwent renovations. The six-storey building's single existing all-gender-inclusive disabled toilet was closed during the initial phase of renovations, so users (including myself) were required to go to a separate building a short distance away, though after hours that building was *also* unavailable and necessitated a call to security to be let in.[7]

[6] Even the more generous interpretations of the available demographics at the time, Delahunt et al. (2018) would estimate the number of nonbinary students at the University of Canterbury at ~77, from a population of 11,000.

[7] After speaking to the university, I was given a special clearance to allow me into that building to use the bathroom, and I specifically did not schedule participants after-hours or during the weekends for those months, as I felt it would be awkward and embarrassing to "escort" participants to the toilet.

FIGURE 4.2: A photo of the pair of inclusive toilets near the interview area (top), as well as a close-up of the gendered graffitti they attracted (bottom).

The second and subsequent stages of the renovations involved converting a pair of toilets from single-gendered (one male, one female) to inclusive. This pair of toilets became a second site of ideological conflict regarding normative gender. The signs on the toilet were first defaced with human figures with exaggerated genitals (See Figure 4.2). These were scrubbed off and also accompanied by further graffiti representing inclusive use (including a popular transgender symbol). Gender icons were then added back to the doors in the form of thick metal stickers matching the existing signage in colour, with a construction adhesive applied. I consider this an escalation in the ideological conflict, as the stickers were not something you would expect someone to have on their person, like the pen and marker graffiti applied earlier. These stickers were likely purchased for the express purpose of applying to the bathroom door, intended to look "official", and being made deliberately difficult to remove. I was unable to determine who was responsible for this second set of stickers, though there are two possibilities: Somebody may have purchased the stickers at their own cost from an industrial signage or hardware store (where they retail for between $10 and $30 NZD each). Alternatively, the stickers may have been applied by the university themselves, either because the toilets were always planned to be gendered, or as response to the complaints and graffiti.

Though I was unable to confirm who was responsible for the stickers, the University of Canterbury itself took the fortunate final decision to enforce the toilets as all-gender inclusive. The binary stickers were removed and the graffiti ceased (as far as I am aware). The "Toilet" signs were themselves were replaced a short time later with updated "Wharepaku" signage, and remain gender inclusive through the end of 2021.

FIGURE 4.3: A second instance of gendered discourse on the toilet signage after the previous graffiti was removed. A metal sticker was added to enforce the toilet as single-gender only, with graffiti added in "response" indicating gender-inclusion via the addition of a relatively popular transgender icon ⚧ (left) and the "+ ♂" (right).

4.4 Data Collection

Audio recordings and selfies were collected from 18 speakers (8 nonbinary, 5 male, and 5 female) between March 2017 and March 2019 at the University of Canterbury in Ōtautahi Christchurch, Aotearoa New Zealand.

4.4.1 Participants

Participants in this study consisted of 16 native speakers of New Zealand English (NZE). Six participants identified as nonbinary,[8] five as binary female, and five as binary male. Nonbinary participants were recruited first. Initially, I intended to match each nonbinary participant to one binary female and one binary male participant, based on demographic information. However, due to the very narrow demographics of nonbinary participants, age became the major determiner of participation for binary participants. Thirteen participants (five nonbinary people, four binary men, and four binary women) were between the ages of 18 and 22, while the remaining three (one each identifying as nonbinary, female, and male) were between the ages of 44 and 52. I refer to these latter three as the "older triad" throughout this work. Table 4.1 lists

[8]Two additional nonbinary participants were recorded but their data are not presented here due to differences in dialect, though I discuss them briefly in the next chapter. Two additional binary men completed the initial paperwork but did not complete the study: One withdrew participation before taking any recordings, and the second completed the In the Wild (ItW) recordings but due to equipment malfunction rendering the recordings unusable, was unable to continue.

each participants' gender group, pseudonym, age at the time of recording, and their self-identified gender as provided verbatim from the demographic questionnaire (found in Appendix B). While not analyzed further within this thesis, it is interesting to note that there is some variation in self-identified gender even among the binary participants.

Group Pseudonym	Age (Y)	Gender (Self-Defined)	ItW (#)	Recorded (HH:MM)	Transcribed (HH:MM)
Nonbinary					
Alex	47	Non-binary?	6	1:31	1:31
Avery	19	Genderfluid	3	1:15	1:15
Istus	20	Femaleish but kind of not words are hard	3	1:11	1:11
Kit	20	Non-binary	6	1:52	1:52
Mello	19	*Participant's Name*	3	0:41	0:41
Ollie	21	Non-binary	6	1:30	1:30
		Totals:	27	8:00	8:00
Binary Women					
Jesse	19	Female	6	1:09	0:28
Kelly	44	Cis female	5	1:21	0:40
Quinn	19	Female	3	0:36	0:17
Reese	18	Female	3	0:41	0:20
Remi	19	Cis female	3	1:23	0:37
		Totals:	20	5:10	2:22
Binary Men					
Devon	19	Male	4	1:20	0:35
Jordon	18	Male	3	0:33	0:13
Noel	52	Male	3	1:03	0:20
Riley	22	Male - straight	5	1:13	0:22
Taylor	19	Male	6	1:27	0:57
		Totals:	21	5:36	2:27
			68	18:46	12:47
Researcher (Jacq)	32	Nonbinary	n/a	17:41	7:48

TABLE 4.1: Participant Demographic Details. *Recorded Duration* refers only to moments within recordings where participants are speaking, rather than the entire length of the recording (including silences, other interlocutors, substantial periods of nonspeech like laughter, etc). *Transcribed Duration* is the subset of the recorded duration that is fully transcribed in ELAN (The Language Archive, 2021). As nonbinary participants were fully transcribed, *Recorded* and *Transcribed Duration* are identical.

4.4.2 Data Collection Procedures

After the initial contact (usually via email or facebook messenger), participants were asked to complete three stages of data collection: An initial meeting and equipment

tutorial, ItW Recordings with selfies, and a Sociolinguistic Interview. These are outlined below.

Initial Meeting

At the initial interview, participants were told the scope of the research and the ethics permissions involved, given the opportunity to ask questions, completed a short demographic questionnaire, and taught to use the equipment. They were asked to record themselves speaking to friends, family, coworkers, and other people they interacted with. The breadth of the recordings was emphasised over quantity. They were also asked to take a "selfie" adjacent to each recording. When being shown how to use the equipment, they were shown how to set the mic recording level, and told the recommended setting was 2 (out of 100), after Gnevsheva (2015). The intention of this low recording level was to minimise the chances that the other interlocutors would be picked up by the mic and audible on the recording. However, every participant was told that there were limitations for the technology, especially in certain environments and with certain speakers. Participants were asked to relay this information to interlocutors and receive consent to record from all parties involved in each recording, but this was handled informally and no data or formal consent paperwork from interlocutors was collected. It was sometimes the case that interlocutor(s) were audible on the recordings.

4.4.3 Materials

All recordings were made on a Zoom H4n recorder, using Beyer Dynamic headset microphones. ItW recordings were made with the participant wearing a single microphone. The exact recording conditions for the ItW recordings are unknown, but contextual information for each recording can be found in Appendix C. The Zoom's mic recording level stayed at 2/100 following Gnevsheva (2015) with some exceptions due to user error in setting the level before recording. All recordings were taken in stereo (single channel) at 44,100Hz (1,411 kbps bit rate) as .wav files. All sociolinguistic interviews were conducted in the same private observation room in the Linguistics department at the University of Canterbury. All interviews, with three exceptions, were conducted using the same Zoom H4n the participant recorded their ItW recordings on, with a second Beyer Dynamic headset microphone used by me. Interview recordings were taken in dual-channel stereo with each mic feeding a single channel, at 44,100Hz (1,411 kbps bit rate) as a wav file. The exceptions to this set-up were: Avery's interview does not contain the researcher on a separate mic channel; Istus and Alex's interviews were recorded with two separate Zoom h4n devices, and; Jesse's ItW recordings were taken in 705 kbps mono due to an accidental change in the recorder's settings.

ItW Recordings

Participants kept the equipment for 1-2 weeks on average. I was available during that time to troubleshoot, but for the most part there were no problems. When the recordings were completed, participants returned the equipment and the final interview was scheduled for a few days later, to give me a chance to listen to the recordings and prepare for the interview. A list of general details of each participant's ItW recordings, including time/location of the recording, number of interlocutors

and their relationship to the participant, and general topic of conversation, can be found in Appendix C.

Selfies

Participants were asked to take a selfie of themselves to accompany each recording. These images were collected because of the inescapable importance of appearance to gender identity. No special instructions were given regarding the selfies, and while some participants focussed on entire "outfits" (photos in front of a mirror or taken by other people), some took selfies of their face only, and most were from the shoulders or torso up. Many images, especially from the final interview day, were not taken or were taken during the interview and not forwarded to me. The selfies proved to be excellent for soliciting unscripted speech and self-reports about physicality from participants, as well as jumping off points for other discussions about style, fashion, gender, and presentation. For the most part, this thesis leaves aside analysis of the selfies on any sort of macro or comparative level, though Chapter 9 examines the relationship between acoustic queues and changes in a single visual factor (use of makeup) across recording contexts for a single nonbinary participant, Istus.

Sociolinguistic Interview

A few days after dropping off the equipment and their recordings, participants sat down with me for a sociolinguistic interview, after the format pioneered by Labov (1968; 1972). Interviews ranged between 30 and 130 minutes. The questions were drafted ahead of time, and while some were adopted directly from the traditional methodology in Tagliamonte (2006, Chapter 3), the others were created with the intention of discussing sex/gender explicitly, while also evoking more and less formal speech in participants. The research context was somewhat helpful in this case, as even relatively mild questions involving sex/gender evoked a response in participants referencing current media and politics. However, it must also be acknowledged that there is a level of ethical reflection involved in deliberately evoking "emotive" speech, particularly in marginalised participants regarding their marginalisation.

Interview questions were broadly organised into blocks (described below), with the interview structure beginning with general identity questions, then alternating between discussing individual ItW recordings and topic blocks, and ending with the meta questions regarding the participant's impressions of the research project. I attempted to create a natural flow to the interview by linking topic blocks to ItW recordings, but not all questions fit naturally into the overall interview structure for all participants. For example, not all participants spoke to family in their ItW recordings, and so there was no natural time to ask the question in the family block of questions. In these situations, questions were asked either when they arose in other parts of the interview (for example if a participant mentioned their family while discussing where they acquired clothes), or were asked a few at a time between ItW recordings. The resulting interviews progress in a relatively naturalistic way across participants recordings, with the ItW recordings themselves providing a loose scaffolding and a kind of "reset" between topics. A complete list of interview questions, organised by topic block, can be found in Appendix B. A brief description and some examples from each topic block is described here:

1. **Opening Questions**: Stage-setting questions about identity and gender, including *"What is your gender?", "How do you feel about being <gender>?" "How can you*

/ Can you tell when someone else is <gender>?" Nonbinary participants were also asked about their personal histories, such as when they first put a name to their gender.[9]

2. **Recording Details**: Detailed questions about each ItW recording, such as "*Where was this recorded?*", "*What was the conversation about?*" Nonbinary participants were also asked *"Does the person you're talking to know you're <gender>?"*.

3. **Selfie Details**: Detailed, open-ended questions about each selfie provided, including *"Tell me about this outfit."*, and questions about specific choices that seem to stand out – such as a bag or hat that showed up in multiple photos, or what the logo on a shirt meant. Over the course of the interviews, I would also ask about style outside of the selfies, such as general use of and attitude towards makeup, glasses, fashion, etc.

4. **Physicality**: Questions about the participants relationship to their bodies, and physicality: *"What is your favourite/least favourite part(s) of your body? Why? Do you do anything to hide/display that part?"* Participants were also asked more broadly about their opinions regarding the relationship between gender and physical appearance.

5. **Interventions**: Nonbinary participants were asked if they had sought medical or nonmedical intervention for altering their gender presentation in ways that may have impacted their speech. I explained before they answered that I was not concerned with treatments that would not impact speech production directly (as per the best practices outlined in Davies and Goldberg (2006, p. 5)). Medical intervention included anything that involved direct consultation with/from health care personnel, such as hormones, surgery, or vocal training. Nonmedical intervention included anything undertaken by the participant without medical supervision, such as binding, tucking, ingestion of hormones or substances thought to contain hormones (such as borrowed birth control pills or soy), etc. If participants identified an intervention that was temporary, they were asked to identify whether the intervention(s) was/were in use during each recording.

   All participants were also asked if there was anything else that might have altered their voice from recording to recording. Among nonbinary participants, Ollie and Kit identified instances where they were binding, and Istus mentioned some changes in diet. One binary participant, Devon, identified recordings where he was wearing elastics on his braces.

6. **Family**: Participants were asked a few questions about their family, including whether their definition of, and relationship to, family had changed over time. Participants who identified as nonbinary or otherwise LGBT+ were asked whether they were "out" to their family. Participants who identified more than one family were asked about each in turn.

7. **Language**: Questions in this block were adapted from Tagliamonte (2006, Chapter 3) (which are themselves an adaptation of Labov (1973)), and asked participants to describe the way they and the people around them talk. Participants were also asked to describe their own voices, in the framing of describing how they sounded to a friend they knew online but had never spoken to. While

[9] Note that <gender> was replaced with the label provided by participants, either in their demographic questionnaire or as they responded to the question "*What is your gender?*"

not given explicit instructions to describe their voices in gendered ways, many participants did so (which makes sense in the context of the interview and research itself). Most participants also described their voices using features of their New Zealand accent.

8. **Network Size**: Nonbinary participants were asked about their network size, either explicitly during the interview or during some previous interaction. It is important to note that many nonbinary participants did not have wide networks of regular contact with other nonbinary speakers, and all participants identified the internet (particularly Tumblr and Youtube) as sources of knowledge/exposure to other nonbinary people and voices. Some participants reported not knowing any other nonbinary people at all (or that I was the first other nonbinary person they had met outside of the internet), and discussed their own language vs their peers, or what they noticed in media (e.g. on youtube). Others talked more generally about LGBT+ or "SJW" (Social Justice Warrior, a tongue-in-cheek name) communities, both groups known to be generally more inclusive/accepting of nonbinary genders than others.

9. **Philosophical**: These questions were intentionally broad, such as *"What is gender?"* and *"What were you told about the difference between boys and girls growing up?"* Participants were also asked to define femininity and masculinity, and about their relationship to each, though unfortunately the presentation meant that these were usually – though not always – answered in tandem, and often in oppositional ways (masculinity is X, femininity is not X).

10. **Meta**: Finally, this section consisted of three questions, two asking the participants for their impression of the research (*"What do you think I'm studying? What do you think this research is about?"*) and the last an open-ended "intersectionality" question. I first highlighted my own positionality in this work (as an outsider / Canadian rather than a Pākehā New Zealander, as a researcher rather than a participant, and as a white nonbinary person). After this example of my own context, I asked participants if there was anything they wanted to discuss or share about the way their personal context and identity intersected with their gender, giving examples such as race, sexuality, religion, or (dis)ability. While this question could be the grounds of an entire interview on its own, I did not engage with it in a satisfactory way, particularly because I did not think to step outside my own context. I discuss this more in Chapter 10.

If a participant seemed particularly interested or attached to a certain topic, I would sometimes ask them to expand or talk about it more in relation to their gender. For example, Ollie really enjoys skateboarding and identifies strongly as a skater. During my interview with Ollie, we talked about skate culture, and I asked him about his relationship to skate fashion and jargon.

4.5 Making a RAINBO

4.5.1 Transcription

ItW and interview recordings were transcribed in Elan (The Language Archive, 2021) using a template modified from Clark et al. (2016). There were separate tiers for Participant, Interviewer, and "Entities", indicating identifying information that should be low-pass filtered before any kind of public export/display. While all recordings

were segmented at the utterance level and marked to separate interview questions from participant responses, the sociolinguistic interviews with binary female and binary male participants were only partially transcribed due to time constraints. There was no specific methodology for how utterance breaks were determined, though single turns were split into multiple utterances if there were long pauses or shifts in topic, as shorter utterances increase the reliability of the automatic alignment. The mean utterance length as measured in Fromont and Hay (2012)'s Language, Brain & Behaviour Corpus Analysis Tool (LaBB-CAT) is 4.3 seconds (19:57:10 or 71,830 seconds total participant duration divided by 16,756 participant utterances). As of writing, RAINBO contains a total of 34:39:17 recorded hours of speech across 17 participants (including the interviewer, me). Excluding the interviewer transcriptions, RAINBO contains 19:57:10 hours of transcribed speech and 146,788 word tokens across 16 participants.

4.5.2 Tagging

Tags were applied to interview recordings to demarcate areas of interest or shared context across responses. It is established by previous work that certain topics (Coupland, 1980) and indeed even certain words (Hay and Foulkes, 2016) and other metacontextual aspects of speech can elicit different productions. In approaching this type of tagging, my intention was to strike a balance between the kind of microtagging required by fine-grained discourse analysis (as seen in Kirtley, 2015), and a broader corpus-style analysis. While I do not delve more deeply into a topic analysis within this thesis, I do explain the process here. Like the selfie data, the topic tagging was undertaken as part of making RAINBO a corpus suitable to future work and expansion.

Three types of topic tagging were applied to each sociolinguistic interview, explicated in more detail below but broadly tagged in terms of topic/subject, person, or temporal space. Tags were applied to the entire response time of a given question, and some tags have multiple time stamps, for example if there was an interaction with the interviewer and then the participant continued to discuss the same topic, or a participant circled back around to an earlier question later in the interview. For the most part, participant responses to questions were direct and relatively short, so while many tags are coded for multiple topics (Family and Gender, for example), no single response contains a complete shift in topic over its duration. Out of 1170 response blocks, about 40 are longer than 2 minutes, and about 230 are longer than one minute.

Topic Tagging

Topic targets fell into broad categories that were similar to the ones used when designing the sociolinguistic interview (enumerated in the previous section). These categories are intentionally broad, as they are meant to intersect with the other topic tags to create a more nuanced picture of the responses. Topics that did not fall into one of these topics was left untagged/Other, though these are rare (for example, there are few untagged utterances of Ollie's where he is discussing the difference between skating and scootering, but many of his other utterances about skating are tagged by topic, such as skate fashion or the slang used by skaters.

1. **Gender**: While the entire context of the interview is one in which gender could be considered a relevant topic, this referred to questions that explicitly asked

about gender, either specifically involving the participant's gender (e.g. *What is your gender? / Does your family know your gender?*) or by direct relationship between the question and broader gender stereotypes (e.g. *What is masculinity? / What were you told about the differences between boys and girls growing up?*).

2. **Family**: Questions about family involved asking the participant to define their family and explicate their current attitudes and relationships towards their family. It also included questions about their family's attitudes, both now and in the past. This tag does not make a distinction between the type of family identified (some participants separated biological family from found family, while others made distinctions between their immediate family and their extended family or in-laws).

3. **Language**: Responses specifically about features of language, including both phonetics (how people sound when they speak) and content (what people say).

4. **Physicality/Fashion**: Responses about physical states of being, bodies, clothing, and fashion. While it is somewhat disingenuous to collapse something relatively stable like "body shape" with clothing and other fashion choices, certain topics – such as why certain clothes might be flattering on certain bodies – are inextricably linked.

5. **Meta**: Responses involving meta discussion of the experiment itself. This is not necessarily an illuminating topic, but was necessary as these types of responses did not fit neatly into other categories.

Target Tagging

These tags refer to the relationship between the participant and the interlocutor. broadly after Coupland and Bishop (2007).

1. **Self**: When the participant is talking about themselves, their own ideas, and states of being.

2. **Close**: A fairly broad category encompassing when the participant is speaking about family or close friends.

3. **Far**: As intended, the "Far" label refers to targets the participant is speaking about who are not members of their family or close friends. In practice, the distinction between "Close" and "Far" tags are not always clear and there are a lot of relationships that may be considered liminal. Rather than trying to create a "Middle" category, when labeling I erred on "Close" except in a few cases where it could be reasonably assumed that the participant and the target(s) did not know each others' names, or where there was a marked difference in authority without asserted friendship (for example, talking about a teacher or boss without qualifying that they are close).

4. **Theoretical**: This tag refers to when the participant is talking about people that do not exist, or nonspecific generalisations of people or a person. (For example, a participant talking about what TERFs in general believe is coded as "Theoretical", while a person discussing J.K. Rowling is coded as "Far".)

Time Tagging

These tags refer to the period in time being discussed

1. **Far past / Childhood**: For nonbinary participants, this is the point before "knowing" they were nonbinary (that is, before finding a label and/or identifying themselves as not belonging to traditional male/female categories).
2. **Near past**: This is a broad category that covers around "a year or two ago" (for most participants, generally the last year of high school and/or first year of university) all the way up to the date of the interview.
3. **Present**: Refers to "right now", including the interview date.
4. **Future/Speculative**: Responses discussing future events or states. This also includes speculative/possible responses that are not always temporally located in the future (for example, "If he had X'd, I might have Y'd").

4.6 Conclusion

This section described the research methodology and data collection process in creating the RAINBO corpus. It also spent some time examining the particular research context in which the data was collected, and provided the participant information from a demographic standpoint. In the next chapter, I take a closer look at each of the sixteen research participants, and take a more personal perspective. I present each participant and introduce them in "their own words" via quotes from their recordings, as well as my personal impressions.

Chapter 5

Meet the Participants

5.1 Introduction

This section gives a brief introduction to each participant, including my impressions of their personality and style, and relevant quotes from their recordings and interview. It is impossible to present the entirety of their research participation in a few paragraphs, and equally impossible for me to know an entire person from a few recordings over a few weeks. I hope that the following pages will serve as a kind of bridge between the three of us – reader, participant, and researcher. I have endeavored to point out the connections I made between myself and each participant, and the ways each participant influenced me and the research overall. Where possible, I have tried to provide quotes that allow each participant to present themselves "in their own words." As the nature of the sociolinguistic interviews is recorded conversation, there are often hesitations, repeated words, or small interjections. For readability in text format, in cases where I do not think it alters the intended message too much, I have edited some of the quotes to remove these aspects of spoken speech in an attempt to improve their readability in text form.

All names are pseudonyms. With one exception[1], I have assigned names to nonbinary participants (following the methods outlined in Drager, 2015) based on personal associations I made between their interviews and personalities, and other people and characters I knew. Binary participants were named by compiling a list of 30 or so names from various websites labelled as "androgynous", "gender neutral", or "unisex", and assigned randomly using random.org.

The analysis throughout this dissertation takes a hybrid approach, incorporating more general qualitative and quantitative methodologies throughout. For each statistical analysis and visualisation, I want to emphasise the individual humans driving these productions, and for the weight of that aggregate data to bolster every Spotlight featuring an individual. In addition to the insight provided by this hybrid approach, there are three more personal reasons for choosing to engage with the research in this way. The first and most primary is my real desire and dedication to allowing space for qualitative interpretations within quantitative analysis and vice versa, though my training and competence is more heavily in the latter. The second is because I am not the kind of person that is comfortable entering a “community” in the traditional ethnographic sense at the best of times, and certainly not as a 30-something immigrant seeking to do research and still interrogating and discovering my own nonbinary identity. Finally, the nonbinary "community" in Ōtautahi Christchurch and beyond is much looser than other ethnographic groups. In fact, none of my participants know each other[2], and nonbinary folks are so “rare” that any community

[1] Alex chose their own pseudonym.

[2] It is possible that they have since met, as the city is relatively small and the number of community gatherings with a focus on nonbinary people is growing, much like the nonbinary population itself.

I could conceivably access would be primarily through Lesbian, Gay, Bi, Trans, and Others (LGBT+) communities foremost and via snowball recruiting, which would make it difficult to separate an emergent "nonbinary" dialect and strategies with a more general community-level one.

While not exactly a direct conflict or contradiction, this hybrid approach must also deal with the tension between methodologies that utilise different levels of feedback and participation between the researchers, the participants, and the work. While variationist approaches tend to favour the objective presentation of work and participant anonymity (being aware of the ways the work may harm communities or individuals), the ethnographic approaches maintain a focus on participant-*driven*, often iterative work, in which researcher positionality is emphasized and the goal is the mutual benefit between community/individuals and the research, as opposed to the extraction of data with the least amount of harm or interference.

I realise there are ways in which this research would benefit from more feedback from participants. While I have endeavored to present them "in their own words", and make clear the places where I am inferring or speculating about something they did not say explicitly, there is no way to remove my own interpretations completely from this work, and in fact it would be misleading to try. However, before speaking to a single participant or hanging the first flyer, I submitted an ethics application that kept the anonymity of the participants paramount. This included removing their personal identifying/contact information from their data, and letting them know the date at which their identifying information was destroyed. As I continued to analyse the data, I realise I am unable to contact participants for more information (such as ensuring they are okay with the pseudonym used), or to allow them to correct any errors or omissions I have presented. This is frustrating, and I do worry that not allowing my participants the opportunity to see the final product to make corrections or elaborate will cause harm. Ultimately, while *all* research involving humans is collaborative in some sense, the ethical methodology I have utilised in this thesis prioritised anonymity over some forms of collaboration.

5.1.1 Another Note about Disclosure

While in a previous section I discuss disclosing participant Assigned Sex at Birth (ASAB), in this section I discuss my own disclosure to participants as a nonbinary person. While I strove to keep all participant experiences “the same”, in that I did not disclose directly that I was nonbinary, I did not try to hide my identity, either, at least not consciously[3]. My facebook profile (from which the recruitment materials were posted) uses they/them/their pronouns for me, as do other people in the department, and my office contained some small “pride” objects: the most overt of these would be a sticker of a dragon holding a nonbinary flag on my laptop. These tend to fall into the realm of covert signaling, in that they signal identity to in-group members (Berger and Ward, 2010, p. 558), but are unlikely to be noticed by others or, if noticed, unlikely to be understood. There were also aspects of my work “out in the world”, so to speak: I had entered and placed in a 3-minute thesis competition at the University of Canterbury, and as a result I had been interviewed for two institutional magazines, one targeted at students and one at alumni. (As mentioned in Chapter 4, my research was taking place after the election of Trump and at the height of a second wave of the so-called bathroom bill discussions around the world, so the topic was relevant

[3]I do feel in one case I was not forthcoming about correcting a participant who made an assumption that I was binary and cishet, technically a lie of omission, but it was during an interview that was already slightly uncomfortable for both of us and I worried it would have thrown it completely off-kilter.

and top-of-mind.) Only one participant, Alex, mentioned any of the more public references to my work, when they pointed out they had signed up for the research after seeing the article in the student magazine.

Recruiting materials also changed throughout the course of the research. Initially, I was looking for binary and nonbinary participants concurrently, but very soon prioritised nonbinary participants before focussing on binary "controls" to match the demographics of nonbinary participants (this ended up being a non-issue due to the relative homogeneity of nonbinary participants, with the exception of the older triad). This decision was motivated by access, as I only had two sets of recording equipment available, and wanted to minimise the risk of having to turn away a nonbinary participant due to the equipment being unavailable for use. This serial approach to recruiting was also due to the vandalism/commentary on the recruiting posts. In the later phase of recruitment I simply removed the "nonbinary" portion of the material, causing a complete drop in outside commentary/engagement with the recruitment posts.

All of this preamble is to note that, despite efforts to "act the same" with all participants, and taking pains to be careful about the information about myself presented, what ultimately happened was still a lopsided experience, whereby I am reasonably certain that every nonbinary participant knew my own identity as nonbinary, while only a handful of binary participants did. Even within binary participants, none directly asked if I was nonbinary, though some did comment on my pronouns or objects in my office that indicated they were aware of my nonbinary identity.

5.2 Participants

5.2.1 Nonbinary Participants

Avery - They/Them

> "I feel like I probably lean towards more the masculine side, but I would still identify as genderfluid."
>
> "I'm not a man. I'm a *dude*. I'm a *boy*. *Boy* and *dude* just sound less gendered to me than *man* because I just can't view myself in that stereotypical masculinity."
>
> "I think being comfortable with having my parents know that I was gay probably helped me to be comfortable in myself with experimenting, I suppose, and just thinking about myself differently because I had always felt like that stereotypical masculine role [was something] I really did **not** fit into. So after I was able to be open about my sexuality, being able to be open about my gender to myself was a lot easier."

Avery was the first interview I conducted. While there were awkward moments, Avery was so outgoing and kind, I felt like they were working to put me at ease and reassure me that they knew what they signed up for, and we would get through this first one together. At one point very early on, I commented that the questions I was asking, which had seemed like such normal "sociolinguistic interview" questions on paper, were a little more creepy when you are actually asking them, sitting alone in a room with another human being. Avery leaned into the joke, dropping into ventricular voicing and saying "tell me everything about your life" and we both

laughed, breaking the ice completely and setting up the tone of the interview; a little bit ironic, a little bit serious, but both of us knowing there were going to be some weird moments to get through.

Avery's initial presentation, their voice and mannerisms, evoke a very particular trope of a gay dude[4] that permeates popular media. Avery talks about emphasising that flamboyance when they feel more comfortable, and about how they have always been a fairly introverted and shy person ("acting like myself is something that I used to probably struggle a bit with"), and how growing up with a sense of belonging, both among friends and within the broader LGBT+ community after coming out as gay, really helped them be more expressive and boisterous. Avery is part of a huge blood family, spread around the country and the world, and they feel closer to many of their friends than their siblings or relatives. They talk about being quiet and reserved when they are around authority figures, or feeling shy or unsure, but when they are comfortable, they say, it all comes out and they can be super "swishy."

Their recordings are a typical slice of life of a Canterbury university student: A conversation with their mum on the way home from a MuSoc show; a conversation with a friend/classmate about uni classes while making drinks; a conversation with an old high school friend while gaming. The topics are a mix of gossip, catching up, and joking, with Avery's honesty and kindness showing through in every one. They pause when asked a question, often making a quick joke before answering more seriously, and often compliment the people they are speaking with. Avery's selfies are well-lit, well-posed, and centred on their face. Their eyebrows are thick and dark, shaped deliberately. Their eyes and eyelashes are accentuated by eyeliner and mascara in most photos. In one, they are holding a hand up as well, showing off dark nail polish. They have multiple facial/ear piercings, the bold makeup juxtaposed with many smaller brightly-coloured and sparkly studs and delicate design work. In every image they have visible facial hair, ranging from five o'clock stubble to a few days' growth. Avery mentions being hairy a few times throughout the interview, and discusses how their use of make-up is one way they counteract the hypermasculine association of body hair, to push their presentation towards something more gender-ambiguous.

The patience Avery extended to me during the awkwardness of the interview is apparent in other parts of their life. They talk about leaving space for their mother to learn ("though she's always accepting, she's also sometimes just a little bit off target"), to take the time to explain things to her when she missteps in the well-meaning ways, such as offhanded/complementary misgendering like referring to Avery as having grown into "a good man".

Avery's experience of nonbinarity is one layered with and coexisting with their identity as a gay dude. They do not have an explicit moment of coming out as nonbinary to their family ("after I've already come out as gay and done all of that excessively awkward stuff [...] just don't want to do it again"), but have brought it up in conversation or more casually – correcting their mother's use of hypermasculine terms like "man", for example, or expressing discomfort with body hair and other gendered norms. For Avery, genderfluidity is about experiencing feelings and identifying with aspects of both binary genders, to different degrees, though they say they lean a little more towards masculine. For them, being Assigned Male at Birth (AMAB) influences the ways they express femininity: "[T]he female side of

[4]Avery expressed multiple times in their interview, including in the quote provided, that they dislike being referred to as a man, so I have taken care to replace this with alternative terms. I do not intend this to come across as infantilising them, but to respect that their masculinity is not the stereotypical kind that terms like *man* and *male* evoke.

things comes through just a lot more swishy, flamboyant, than it would with the roles reversed." For a time in the past, before they had begun identifying as genderfluid, they were shaving everything off, and going to the gym often ("I didn't want to be muscly I just wanted to look like a stick"), and being unhappy. But they talk about embracing the masculine side of themselves more, as time passed, "As I've gotten older I've kind of become more comfortable with the masculine side of things [...] I feel like [identifying/being genderfluid] also made me more comfortable just in any form of expression."

Istus - She/Her

> "Deep down I'm this, you know, high-femme, except that I'm a weird genderless potato person."
>
> "Generally I try and present like the least gendered way possible, um, and then some days I'll have like the high femme day where it's dresses and makeup and heels and pretty sparkles and it's *winter* [so] I don't want to do any of that, it's *cold*."
>
> "If I'm presenting like really female that day I'll generally try and talk less."
>
> "Most cisgender (or *cis*) people tend to take everything at face value and stuff... literally all I need to do is put on a dress and a bra and I get read as female, but if I do anything short of that it's like, "*Oh, you're a man*.""

Istus came to the final interview in a three piece suit and lipstick. The impression I get from Istus is that she uses self-deprecation like a shield. She jokes about the things that hurt her deeply in a way that would be jarring if it were not a pretty familiar brand of LGBT+ humour. She is secure in her identity but also incredibly vocal about the struggle in making it work for her, and the disconnect between wanting to be "high femme" and the unrealistic standards of femininity, and what she feels is the way she presents most of the time, as a "genderless potato person".

Istus is funny, with the chaotic energy of the zoomer/netkid raised online with memes and YouTube. She makes jokes about the newest viral memes unprompted. When asked how she can tell if someone is nonbinary, she points out that we are much funnier than cis people, presented as a matter of objective fact. As one piece of evidence, the says that binary people would never refer to heaven as the sunchips place, or have profile pictures of Griffon McElroy[5]. At that point I asked if she had seen the facebook recruitment, as my profile at the time used a Griffon McElroy profile picture, but it turned out to be an unintentional callout. Istus' pseudonym is a reference to this same injoke.

Beyond making jokes about the internet, Istus' humour and attitude is based in keen observation. Istus has a "stealth femme" outfit that she points out looks neutral/masculine when worn, but is made up entirely of women's clothing, either by brand, fit, or specific convention (for example, the side the buttons/zipper is on):

[5]Both of these are references to the McElroy brothers, a trio of cishet adult brothers and their extended families whose pantheon of podcasts and live shows were pervasive throughout the internet in 2018, especially among LGBT+ and otherwise marginalised communities, for being good allies and their ability to foster parasocial relationships with fans.

> "I have this specific outfit which is, everything in it is a designatedly female item. But, [you] can't tell from the resulting outfit. [...][t]his flannel shirt buttons that way and these are women's skinny jeans and, you know, these shoes have much more heel in them than you'd expect from a male shoe, and stuff. And you can't like, you can't look at the resulting outfit and go "This is female presentation." – *But **I** can*."

Istus' interview is deeply personal but also non-emotional. I am left feeling like I know Istus, and that we could be friends, but also that I am not sure I have seen her "real" feelings under the jokes. At one point, she made a throwaway comment about not being out to a family member due to transphobic behavior, but this is framed in a lighthearted way, as if she is withholding a secret from a naughty child. She talks about how unobservant people are, about how they could not possibly understand nonbinary identity because the moment they see a dress they slot you into the "female" box, but also about how she tries not to talk in some situations where her voice makes her self-conscious. Istus discusses dysphoria and anatomy with the same frankness and humour as she discusses everything else. She explains the details of tucking[6] to me with the same explanatory tone as she talks about ancient milling processes in her recordings.

Her recordings are all similar to each other, and very different from the recordings provided by other participants. I am presented with three recordings that could be informal Ted Talks: One is about a class assignment, and two are about cooking. None of these three things take the turn you would expect from that explanation. The first involves live animals but mostly about creatively tracking down and building necessary equipment, the second two are about (in one case) recreating an ancient recipe using ancient measurements, and (in the second) the long drawn-out process of preparing a rare fruit for consumption. In all of them, Istus is in the role of a lecturer and storyteller, and she is purely in her element. Jokes abound (including a reference to the ancients heading to *Ye Olde Pack & Save* for some ingredients which still makes me laugh), including the cutting insights that punctuated our interview. The esoteric, sometimes bizarre nature of Istus' recordings emphasises her full commitment to performance – for example, her explanation to a friend regarding her attempt to recreate an ancient recipe, is presented both with the storytelling flair of a high adventure, and also as a guidance for the future, as if this is a perfectly understandable thing for anyone to do on a bored weekend. She speaks enthusiastically, when she is excited about something her words tumble over each other, punctuated by checking in with "you know?", reduced often to just a two-tone rise in intonation *"y'ow?"*

Like most of the nonbinary participants, though, she also talks about the exhaustion of seeking services, of changes she might make regarding hormones or surgery. Istus had the most numerous and specific ideas about the gendered perception of fashion and voice, and detailed her vocal strategies for presenting her gender. While still couched in humour ("the makeup hides the bags under my eyes") and jokes, her frustration is palpable. I do not want to say Istus' presentation is calculated, because that has a negative connotation, and there is no real intent to deceive people on her part. But she is keenly aware of the way she is perceived, and also aware of the connotations of what she wears and what she does. She makes a distinction between things she does for herself, like her stealth femme outfit, and things she does for others, like certain styles of makeup where her intention is for other people to notice she is wearing makeup. When asked about her hair, she delivered a detailed, practical explanation as to why long hair was preferable to sunscreen for protecting

[6] The process whereby someone arranges their penis and testicles to minimise their appearance.

her ears from the sun. She then immediately qualified that explanation as the reason she gives other people, but that its not the *real* reason she keeps her hair long, with the obvious unsaid implication being because long hair is feminine. The selfies she presents mostly focus on her face and shirt, her makeup and the way she has done her hair.

I do not think Istus' multiple reasons for doing gender and the incredible amount of thought she has put into these decisions are exactly unique. But what I *do* think is unique is the frank and playful way Istus is able to identify these reasons, and separate them in terms of identity, presentation, and perception. Istus' insight and ability to articulate her insights has been hugely beneficial for this thesis, and I return to these insights multiple times.

Alex - They/Them

> "Most of my experience with nonbinary people has been on the internet and there's a very standard look and presentation, and this is something I'm still struggling with myself, because I don't want to do that."
>
> "I didn't know there was any other option so therefore I just went, "oh well, I'm female but I just redefine what "female" is to be what I am.""

Alex is one of three participants over 40, who form what I refer to as the "older triad", along with Noel (a binary man) and Kelly (a binary woman). There is a certain kind of stealth afforded to Alex due to their age, an invisibility that is sometimes, though not always, negative. They have a few shirts that "out" them as nonbinary, one listing their pronouns, and one with a more covert signal, in that it requires some knowledge outside of LGBT+ awareness to understand the wordplay involved. The latter is, like the small stickers around my office, only really explicit if you both know what you are looking for and are a member of both groups (Smaldino and Turner, 2020, p. 6). The former, however, is plain to people with minimal knowledge of nonbinary identities (and the importance of pronouns that are attached to them) but much more obtuse to older people. This works in Alex's favour, and they comment about wearing it around older conservative people, that the only people who would understand it as a gender signal would also be the most likely to be accepting of nonbinary identities ("No one's actually gonna know what [my shirt] means anyway there, and if they do, well, they probably *actually* know what it means and that's cool.").

Alex went above and beyond in regards to the recordings provided, ticking every box in the suggested environments, including at home, at work, while out with friends, as well as the only In the Wild (ItW) recording in conversation with children. They are professional and enthusiastic, deeply into data collection and research, and interested in what is happening around the university. As far as I know, they are the only person to have heard about my work before the interview. Their age and life experience manifests throughout the interview. Alex's social circles, like the other two participants that form the mature triad, vary in age much more widely than that of the younger participants[7]. These include casual friends close in age, as well as a wide range of working-age people through work and various social clubs, and much younger people and children (while Alex has no children themselves, their siblings and friends have children).

[7]Recordings from all other participants, who are 18-21 years old, overwhelmingly skew towards conversations with age peers and parents exclusively.

Alex, more than the other nonbinary participants, has a kind of excitement about being nonbinary, about finally having a label that fits them better than what came before. (“I’ve always had this fascination with trans identities, but I also knew that wasn’t me. I definitely wasn’t a male, but there was just this attraction... and I just went *"oh! there’s other options I don’t have to be just female or male"* and, yeah, it just sort of clicked into my head.”). Having this time to try and reconcile their concept of self with a gender that did not fit has given them perspective into, and patience with, others who do not quite understand the concept of nonbinarity. But this has come with more years of discomfort with being slotted into binary boxes, more chafing, which has given them an enormous appreciation and huge sense of relief to finally feel more comfortable in their identity.

Alex’s relationship to their body is tempered by their experiences as well. They frame feelings of dysphoria in wry understatement (“I’m not too keen on my chest”, for example). They discuss surgery as something they have thought of, but decided against as the wait lists in NZ are long and they do not want to bump someone else off or down the list. Alex expresses a lot of frustration with the physical aspects of gender presentation, including both the limited physical templates and models for "nonbinary identity" (“most of my experience with nonbinary people has been on the internet, and there’s a very standard look and presentation and this is something I’m still struggling with myself because I don’t want to do that.”), as well as the frustration of wanting their nonbinary identity acknowledged without having to jump through physical, sometimes physically impossible, hoops to match those stereotypes ("I’m forty-seven, I don’t need to change how I look *now*, I just want to be me but have people know this is not female, this is something else."). Ultimately, in the way they previously redefined femininity to include their nonbinary self, Alex takes similar comfort in the messaging of of the body positivity movement. They make an effort to accept that they exist in a body that might not be ideal, but is ultimately still Alex, whatever it looks like (“I’m just gonna look as good as I can in the body I’ve got rather than try and hide the bits I don’t like”).

Alex is the only participant I continued to contact after the data gathering phase of my research was over. While I did not seek them out specifically, we share a few hobbies and friend circles, and a friendship grew quite naturally after I finished my participant recordings and interviews (and subsequently stopped deliberately avoiding them). The Alex I know is a progression from the snapshot of Alex presented in this thesis. We have spoken a few times about the ways their participation in the research, and our subsequent friendship, has shaped their own experiences as well. Alex is the only participant who has chosen their own pseudonym, and the only participant who has been given an opportunity to read and respond/correct my presentation of them (though they chose not to). I have the luxury of an entire thesis chapter to pontificate on the ways these 18 people have changed my way of thinking about gender. Of all of them, only Alex has the opportunity to tell me what I got wrong, and how the experienced changed them.

Mello - They/Them

> **Jacq: Does your mum know that you’re nonbinary?**
> "I mean, she doesn’t even know I’m vegetarian."
>
> "It’s more normal if you don’t have to announce [your gender]… You don’t go out and say “Hi mum! I’m normal and straight!” "

> "This [discussing hair] isn't even gender anything, this is just [because] my life's a mess! Help!" *laughs*

When asked about their gender, they simply use their own name – They are Mello, and that sums it up[8]. Mello describes their early experiences with gender as one where they never thought about boys and girls as separate. They were comfortable in any situation, hanging out with any group, and using any toilet. ("I guess that can just be kids being kids but I never really stopped… [T]here was never really a moment, I just never settled."). It was not until they got older and started having certain expectations put on them, such as the uniform expectations at a segregated high school, that they started to question their own identity. ("I was like "huh, everyone else seems pretty good about this, why am I…"). They describe what they call a "googling rampage" near the end of highschool, finding a bunch of labels – transgender, tomboy, bisexual – none of which seem to fit. Finally, they settled on just being Mello and rolling with it ("yeah I can be like *this* one day I can be like *that*. It's still me.").

Mello is incredibly active, and does a high-risk, high-intensity sport that requires them to be in peak physical condition. The sport is relatively popular, but participation at Mello's level is much rarer, and the community is relatively small, so I avoid naming it to help maintain their anonymity. Dedicated practitioners like Mello often travel to specific international destinations and tend to adopt (or be associated) with particular, close-knit lifestyles. Mello's recordings and interview are peppered with sport-specific jargon, and I often ask what certain terms mean, only to be greeted with a definition that involves three other words I have never heard before. Sport touches almost every aspect of Mello's life, from their choice in clothing to their attitude towards physical appearance and performance. Following from that, Mello is comfortable with their body in a way I have rarely seen before, that stems partially from their dedication to athletics: They need to trust their body to keep them alive, and that involves knowing its exact limitations ("You have to consider yourself, the body you're performing with.")

Mello's jokes have a swift, brazen quality, often self-directed, that still makes me laugh when I am listening to the interview for the *n*th time. When talking about scouring the google search results for a gender term, they add sardonically, "But you know, Google also told me I had cancer and all of that too." They discuss any discomfort or judgement obliquely, trailing off, making a joke, or shifting the subject: When I asked them if their mum knows they are nonbinary; "She doesn't even know I'm vegetarian." Their work uniform allows almost no room for personal expression, but "when I have that [top button] done up [and] I notice one of the ladies coming through the door I'll be like ***BOOM*** *I'm out come on*." The impression I get from Mello is that they do not really *want* to care about gender, but if the world is going to *make* them care, than the world better be prepared to deal with the consequences ("I may go out of my way to... come across really girly just to prove a point."). Their three selfies reflect this blend of capability and desire to shock: One is them dressed in fuzzy, hooded pyjamas, sticking their tongue out in mock outrage. In another, they are on their way to work, professional and uniformed, smiling hugely. The third has been taken by a friend, with Mello posed fairly lewdly in the middle of their training, looking at the camera with mock sexuality.

[8] Mello's pronouns did not come up in the interview at all. While I have the impression they have no preference at all, I am defaulting to they/them/theirs.

Flynn - They/Them & Keegan - He/Him

I present here two nonbinary participants, Keegan and Flynn, who participated in the research but whose data was not included in the following analyses. Both Keegan and Flynn have voices that deviated from the standard (Pākehā) NZ English accent of all the other participants, which makes quantitative comparisons difficult. In addition, given the relatively small size of Ōtautahi Christchurch, I was not confident I could present their information clearly while upholding the standards of anonymity I promised them. For these reasons, I decided not to incorporate the recordings they provided into this thesis, though they are still very much a part of RAINBO. I discuss this reasoning in more detail, and reflect on some ways I could have approached it differently, in Chapter 10.

Flynn describes their gender and the realisation that they were nonbinary in a way that is very familiar to me, as a bit of a comic tragedy. They talk about the crisis-of-conscience that happens when they really begin to realise their ASAB does not fit. "What if I'm trans?!" they recall crying to an ex-girlfriend[9], "What if I'm not gay and I'm just a guy that wants to be with women!?" This story, describing a very real episode of emotional distress, is presented in a comic way, the tragedy both overblown and reduced, in a way I think many LGBT+ people can empathise with. Flynn's experiences growing up in a multicultural family, and living outside of NZ for many years, were invaluable in pointing out the ways my own cultural context shaped the thesis questions and approach. They discussed the way gender roles and expectations, and the concepts of femininity and masculinity differ across languages and cultures (between their non-NZ and NZ peers, for example), and how that influences their own relationship to gender and its fluidity across contexts.

Keegan loves architecture and board games. One of his ItW recordings is unique among all participants in being a monologue[10], where he describes a three story house he has designed, describing in detail the building materials and layout: Sweeping open staircases and high ceilings, the placement of upper floor skylights to bring in light, and considerations of keeping the people inside warm in winter and cool in summer. Keegan is close to his family, and many of his friends are older than he is, family friends rather than classmates or peers. In addition to being transmasculine and nonbinary, Keegan is a member of other groups that experience stigmatisation and has resulted in some difficulty accessing local community groups for young LGBT+ people. Out of all the participants, Keegan is probably the prototypically "coolest". His selfies are relaxed, from multiple angles, in most of them sporting a half-smile and a slouch. At the interview, he is friendly and laconic, unfazed by most of the questions.

The removal of Flynn and Keegan's data from the analysis, has resulted in a somewhat artificial/limited presentation of nonbinarity in Ōtautahi Christchurch, Aotearoa. I sincerely believe in the need to represent and understand that nonbinarity can take many forms, and that the intersectional aspects of culture, ability, race, and other identities are incredibly important to understanding gender identity. While their participation is not directly visible in the visualisations or statistical analyses

[9] I am not sure if the "ex-girlfriend" in this story was an ex at the time the story takes place, or if she was a then-current girlfriend, now-an-ex girlfriend, but we were both laughing so hard I forgot to clarify.

[10] This was presented as an option in the materials given/explained to participants, but Keegan was the only one to provide one (though some of Istus' recordings come close to monologues, there were still other people in the room during her recordings).

that follow, Keegan and Flynn have both unequivocally shaped this thesis (and, I hope, future work) for the better. I am grateful to them.

Ollie - He/Him

None of the shorter, handful-of-words descriptions really sum up Ollie's nonbinary identity in a satisfying way. He identifies as agender and gay, and is in a relationship with someone who is a lesbian. He describes himself among LGBT+ groups as probably perceived as a gay guy when he is in that context, that he comes across as flamboyant ("more like the masculine gay more than the lesbian kind of gay"). In cishet groups people more often see him as cishet ("more of a just straight masculine cut kind of guy"). Ollie is agender, uses he/him pronouns, and does not care if most of the world (outside his close friends and family) refers to him as a man. ("I don't care as much like cos I know that in my head I'm not [binary]. It doesn't matter.")

In a lot of ways, Ollie's experiences with gender are the closest to my own, which makes the interview easier in a way the others were not. But it also makes it a bit more difficult to disentangle my inferences about Ollie's motivations with what he has actually said explicitly. We bond over a mutual hatred of antagonistic architecture, of belonging to hobbies where homophobia is part of the group culture, of realising coming out the first time was not enough, the exhaustion of needing to sit down with friends and family with a "but wait, there's more!"

During each interview with participants, I attempted to be explicit about the topic of discussion, clarifying what may be left unsaid for the benefit of future transcription. There are a handful of moments in my interview with Ollie where I was sure this had already happened, and we were obviously discussing passing, or navigating relationships with people that predate "coming out." However, going back over the transcriptions, neither of us explicitly name these topics. At the time, and on re-listening, it does seem like our experiences overlapped enough that it was unnecessary to clarify the topic, to a much greater degree than many other participants. However, it is unfair of me to prescribe motivations to Ollie that could very well be my own projections, and I will try to let Ollie's words speak for themselves as much as possible.

At a few points throughout the interview, Ollie reaches out to me, encouraging me to visit a local meetup, to try skateboarding, to participate in the community. I have taken him up on some of these offers, though I actually have not met him again. At the end of the last session, after the recording has been turned off, I vividly remember Ollie asking me if I was actually going to *do* anything with this data. It was not really an admonishment, or a plea, or even presented completely seriously, but it was absolutely voicing the very real frustration behind having given a lot of his time and self towards something that might never come to fruition. I told him I would do my best. During the literal years after that interview, at the times during writing this thesis that I considered giving up, I would think about that exchange. This thesis exists at all, in part, because of Ollie, and the sentiment behind his question.

> "Almost any question can be answered with no more information than was contained in it. When the speaker does give more, it is a gift, drawn from some general fund of good will that is held in trust by himself and the field worker. A deep knowledge implies a deep interest, and in payment for that interest the speaker may give more than anyone has a right to expect." (Labov, 1972, pp. 115–116)

Skate culture is huge in Ollie's life. His clothes are skate clothes, his friends are skaters, he is a skater through and through. Top dysphoria[11] also plays a major role in Ollie's life. He admits to binding[12] much more often, and for much longer, than is medically recommended, practically every time he goes outside. ("If I'm binding it's always masculine. I don't want to be read as feminine.") Ollie describes skating as his escape, his "get out of my head kind of situation." Skating is an intense workout, and one where it is difficult-to-impossible to bind while participating in. Because of this inability to bind and still skate comfortably, Ollie will visit the skatepark in its off-hours, very early in the morning, when other people who might recognise him are not around. ("I want to bind when I skate but I can't so that's why I don't go to the skate park when there's anyone else there"). For Ollie, skating in a safe way that allows him the full range of motion and breathing the sport requires, also means risking being seen, and potentially outed, by onlookers.

All but one of Ollie's selfies are in front of a full-length mirror, his phone in one hand and his other hand in his pocket. The last one has been taken by someone else, so, naturally, both of Ollie's hands are in his pockets. Ollie's uniform, as such, consists of board shorts or khakis, skate shoes, skate-branded tops/jackets, and a baseball cap. That is not to say he is wearing the same clothes in every picture, or that the presentation is without personality. In one photo, he is in a child's room (his sibling's), wearing a fuzzy orange bag with a giant plush duck head on it. In another, a rainbow pride flag is just visible in the side of the frame. This reflects Ollie's personality: Frank, straightforward, and unassuming, but not completely hidden or closed, if you know what to look for or what he wants you to see.

Like all of the nonbinary participants, Ollie's identity makes sense to him, even if the language does not provide the right words to explain it easily. He is comfortable with who he is, and knows he is nonbinary regardless of how the outside world genders him. However, Ollie also talks about what he refers to as two separate lives, one where he is a cis guy and one where he is agender, and he is scared of what might happen if they interact ("I'm really scared about crossing those two lives" and "I guess [I] feel like I'm living separate lives all the time"). Most of his family uses she/her, while his uni and skate friends use he/him. Ollie's ItW recordings are split between these two lives, family and friends.

Kit - They/Them

> "If girls were red, boys were blue, [and] genderfluid people are purple… I'm yellow. It's something completely different."
>
> "I've found from a lot of social media things, that people will say, "Nonbinary isn't a proper gender! It's just people making it up!" And I'm like, if you felt as shit as some people do on some days you would *not* say that!"

Kit has an infectious hyperactivity, an excitement about the things they love that draws me in. They talk to me as if they have known me for years, and we settle quickly into answering questions, barely pausing for the entire (102 minute) interview. Kit

[11] Dysphoria for some trans folks is categorised into *top* and *bottom*: Bottom generally refers to the genitalia, while top refers to the chest area. This short-hand is often used when discussing surgical interventions (e.g. "top surgery" vs "bottom surgery"). Different trans people feel different amounts of top/bottom dysphoria (or none at all), and top/bottom dysphoria is not the only kind of dysphoria there is.

[12] The process whereby someone uses tight-fitting clothes, elastic bandages, or a specially-designed article of clothing to minimize the appearance of breasts.

describes themselves as feeling tired during every ItW recording, though in five of the six I hear the same intense energy in their speech that they bring to the interview. The exception is one recorded in the evening, with their mother, the two of them speaking quietly about schoolwork and cosplay[13] plans. This claim of fatigue surprises me at first, but they mention a few times in passing that they find being surrounded by a lot of people, especially strangers, tiring. Of the remaining five recordings, four are in public or semi-public venues (a restaurant, at university), and/or with strangers, and the fifth is immediately after coming home from a large social gathering. I think Kit puts a lot of effort into making conversations seem effortless for the other people involved, and this explains the slight mismatch between their reported mood during each recording, and the energy I hear in their voice.

Kit is a big fan of anime and cosplay, and it was through cosplay and other cosplayers that they realised they were nonbinary. While Kit does remember feeling bad about themselves during puberty ("I felt really uncomfortable in everything I wore [...] I can look back and pinpoint times where I've sat there and gone "I don't feel like myself anymore."), it was distress while cosplaying a character whose gender matched Kit's ASAB that led a trans friend of theirs to suggest they might be experiencing dysphoria. Looking up these terms and learning more about genders beyond the binary, and with unwavering support of friends that were willing to let Kit experiment with new names and pronouns, allowed Kit to put their past and present into a new perspective. ("Looking back there's a whole bunch of things that make a lot more sense. I didn't know that I was nonbinary as like the *word* but I knew I was something other than [that].") Cosplay remains an important part of Kit's life and their relationship to their gender ("cosplay is a sort of an escape, it's a place where people can feel comfortable. [...] A lot of friends who I met through [cosplay] are transgender or nonbinary or genderfluid."). As a result of this hobby, Kit has the widest network of nonbinary peers, relative to the other participants.

Every selfie of Kit's has some aspect of fandom in it, from a graphic tee for a fictional sports team, to an accessory such as a hat or scarf with a small logo and matching colours, down to "regular", non-fandom branded boots or clothing that Kit either purchased for cosplay purposes, or because it reminded them of a character they like. While primarily focussed on cosplay, and possibly because of it, Kit is very fashion-consciousness. They have a keen eye for clothing and accessories that emphasise the parts of themselves they like drawing attention to, like their hands and their eyes. Kit says they used to hate their eyes, but now appreciate them ("I thought that [my eyes] were too big, that I have like bug eyes, [but] then I got a lot of compliments... [someone] said I have an anime character face with the tiny flicky nose and massive eyes and that was both funny and really nice."). Kit is uncomfortable with their chest and sometimes binds, but finds it exacerbates some of their health issues and is unable to do it frequently or for long periods of time. Though they do like the way it looks ("it makes me feel more comfortable looking in the mirror"), they generally only bind on special days a few times a month, like on the day of our interview, or when cosplaying as a male character.

Kit describes their relationship with binding, surgery, and other medical interventions in a way similar to that many of the other nonbinary participants, as a desire to act mitigated by personal failing or laziness. Kit in particular expands on this thought, "I am very lax, and very lazy, so if it's gonna cause me pain or if it's gonna cause me stress or anything like that, I'm more likely not to do it." Avoiding something

[13] A portmanteau of the words "costume" and "play", referring to the act of dressing up as favourite characters from anime and other media, and attending conventions.

that may be painful or stressful is not really a definition of "laziness" I think most of us ascribe to, but there is a tendency among the nonbinary participants to speak negatively in this way and hold themselves to difficult standards, regarding aspects of themselves or their actions that are not "nonbinary enough". Kit also expresses that more permanent medical options (like top surgery) are something they have only thought about briefly, and they fear making a decision they may regret in the future.

Being unsure about permanently altering one's physical appearance is a source of anxiety for many nonbinary participants. Additionally, many talk (some obliquely, some directly) about feeling pressure, both internally and from external narratives, about what it means to act "trans enough" or "nonbinary enough", and how much and how consistently one needs to act in ways that distance them from their ASAB, to be taken seriously or cross some threshold of validity. I think this conflict is often internalised and vocalised by many participants as laziness, or indecision. *If you were **actually** trans/nonbinary*, the thinking goes, *then you would do something about it*, as if the extensive list provided by the nonbinary participants in this research were not enough (such as clothing and diet choices, gym visits, shaving and binding, name changing, etc). It is worth noting here that many of the binary women talk about their relationship with makeup in similar terms as well, though none make the same direct association between the "laziness" of not wearing make-up as fundamentally altering or diminishing their femininity.

5.2.2 Binary Women

Remi - She/Her

Remi identifies as a cis female "for now." In regards to femininity, she acknowledges a feeling of pride in identifying as a woman, but is currently in a process of conscious introspection, as she learns more about social justice and psychology. ("Do I need femininity?") When it comes to stereotypes, harmful language, and what it means to be a woman, she is in the process of actively retraining herself to recognise that the social cues that signal femininity are taught, but that they are neither exclusive to women ("you don't have to be a woman to wear dresses", "men can be feminine"), nor static and absolute: "The first thought is what you're taught, the second is what matters." Remi is close friends with a nonbinary person who uses they/them pronouns. She acknowledges that there are aspects of nonbinary gender that she "viscerally" does not understand as a cis person, but she is supportive of people being able to find new labels they are happy with.

Remi is one of the most fashion-conscious of the research participants. She comes to the interview in a striking Tangzhuang-style top, navy with frog fasteners and delicate embroidered flowers. A handful of her selfies contain similar tops, and Remi informs me they are vintage, over 30 years old. She liberated them from her mother, who hates the style. ("I kind of wear them out of spite" she laughs, only half-serious.) Her process of deciding what clothes to wear follows a system. First, she checks the weather and temperature, which determines the kinds of shoes and clothes that are appropriate. Once she knows which shoes to wear, she picks the outfit – only some outfits work with some shoes, and she dislikes repeating outfits too often[14]. Finally, she picks her jewelry to match. Remi's style is distinct, she knows what she loves (vintage and classically-inspired designs) and wants to wear things that make her

[14] I am aware this description sounds like an alien describing a normal process. Clothing and fashion has always been something I personally found unknowable and obtuse, and this has made discussing these things with some participants hilariously awkward.

happy. Though she does acknowledge that she tries to make each outfit cohesive from an outsider's perspective, the style and the way the outfits are put together is completely for herself.

Like her choice of clothing, Remi is deliberate in her choice of language as well. She make an active effort to use gender-neutral language and replace problematic words with more "conscious language." She says she makes no connection between gender and physical appearance, but self-corrects immediately to clarify that she probably has a subconscious connection, but does not act on it. She thinks for a moment, and acknowledges that she does not go into men's stores and that she buys women's clothing, too.

Remi's recordings provide another example of her conscious approach. In one, she is participating in a uni club meeting, and offering constructive suggestions on the logistics surrounding an upcoming event. In others, she is talking to friends and family, and at various points talks about her classes, her diet, and her health. She outlines carefully the options available to her (whether between choices of meal or classes), how she has approached each problem, and how she came to a resolution for each, such as class clashes or trying a restricted diet for health reasons while maintaining a budget. Remi is not the kind of person to make a rash decision. She explores all of her options in a situation, and decides the best course of action for herself. She is incredibly self-reliant, and knows very much what she wants and what she likes, but that is not to say she is stubborn. Even during our interview, there are often moments where she pauses to think before answering, then provides both her initial thought, as well as the follow-up.

Remi has dealt with some health issues for most of her life, and this is reflected in the interview. When talking about her body and self-image, she makes a clear distinction between the parts that are failing (her knees, "I have joint pain and I'm 19") and the parts that do not meet the standards of beauty (her thighs, "I've always been told fat is ugly, so..."). For Remi, femininity is encompassed by the ballerina: "They may look pretty and delicate and floaty but they are putting *so much effort* into being effortless." Femininity is feelings of fragility and delicacy, flowers and lace and softer things, but importantly for Remi, does not have to be conducted by a woman.

Kelly - She/her

> "You can get away with anything if you look like you've done it on purpose."
>
> "When you're a woman, if you want to wear jewelry or wear a fancy frock or paint colours on your face you can without sort of stepping over too many social bounds, so I feel my female identity sort of gives me the freedom to play with things."
>
> **Jacq: What have you told your kids about the differences between boys and girls?**
>
> "[Be] aware that women have experiences that [men] don't have. If their path is easier because they read as male, they should be aware that they're not putting road blocks in front of people who aren't."

The first thing I notice about Kelly is her shock of brightly-coloured hair and her magnificent cat-eye glasses. The second thing I notice, as we sit down and get into the interview, is the careful way she speaks, extending the last sound in most words

and enunciating clearly. She points this out herself later on, both in terms of specifics ("[in recordings] I always sound like I'm talking more slowly than I think I do..."), and in the way she thinks it may come across, ("[like] I'm explaining something to a particularly dim two-year-old."). While I do not feel like a two-year-old, exactly, a handful of times throughout the interview I do feel a little naive. Kelly is another member of the "older triad", with Noel and Alex. She is incredibly intelligent, with a wry sense of humour and a keen understanding of human behavior threading through her anecdotes and answers.

In Kelly's recordings, she takes many roles: a conversation facilitator and context-provider between her children and their grandparents; tech support for an older friend; art enthusiast and critic; errand-doer and party-planner; and, in one, what she describes as the wonderfully nice feeling of being present with someone who is coming out of the other side of a bad situation "being less of a therapy friend and more of a *friend* friend." I get the feeling Kelly is a therapy friend for many people in her life. She describes her relationships to people in ways I find incredibly appealing: The aforementioned "friend friend" vs "therapy friend", or describing childhood/university friends as "we have more history together than present together", or being "lovingly frustrated" or "performatively formal" with those she is not necessarily as close to or politically aligned with.

Kelly's personality manifests in more than just her hair colour. Her selfies show off three different kinds of glasses, all fairly different, but all bold, either in terms of colour, shape, or pattern. She favours multiple layers, in blues and teals and soft pinks, colours that compliment her hair, and one jacket that contains all three ("A recent happy op shop find.") Kelly's hair is incredibly important to her. She says it took her about 15 years to work up the courage to finally dye it. In her own words: "[It's] not necessarily trying to be pretty or proper, but actually deciding I can have fun with the way I look, and it does not have to fit in with anyone else's idea of what a lady of my age ought to look like." Kelly used to keep her hair long and spent a long time "wanting to be prettier", but her hair now, short and brightly coloured, is about "not trying to be anything other than what feels right to me right now." Her relationship with her body is relatively positive ("I'm mostly fairly at peace with how I look"), though she is frustrated with some functional aspects: She would like to be stronger physically, and she has a lot of pain that limits her in terms of fashion ("while [pain] isn't an aesthetic thing in itself, it has a big aesthetic effect").

Her identity as a woman, and her heterosexuality, are inherent parts of Kelly's identity, but not ones she thinks about too often. It's not that it is not important, she says, but that it does not preoccupy her thoughts or weigh her down. ("I'm me first, and female second.") Kelly grew up in the era of second-wave feminism, where the manifesto that "girls can do anything!" also came with a subtext of "maybe girls are just a little bit nicer, smarter. Maybe if girls were in charge of the world it would be a bit *better*." However, getting older and having kids that do not fit neatly into gender stereotypes themselves has given Kelly a more nuanced understanding. She now sees those differences as a product of socialisation rather than something inherent, for the most part. To her own children in regards to gender, she has told them that, beyond the physical aspects of their bodies, "they shouldn't feel bound by what's expected of them, and should do what feels right for them."

Quinn - She/Her

> "It gives me a sense of achievement that I know multiple languages and helps me feel more comfortable in my identity."

> "I don't really mind what I wear but I do like to look presentable."
>
> "I'm not a fan of makeup because it takes a long time to put on and take off, and it's not good for my skin, and I don't really need it. So I'm just like, why would I wear it?"

Quinn, like Flynn and Jordon, is multilingual[15]. While Jordon and Flynn report a difference in the way their gender manifests in speech between English and their other languages, for Quinn her inherent sense of femininity is boosted and enhanced by her multilingualism. Especially in an Aotearoa New Zealand context, Quinn considers fluency in a non-endemic language (that is, outside of English and Māori) a marker of intelligence and status ("The more languages a person knows, they're often perceived to be more intelligent, especially so if they're a female."). Quinn is also more aware of her femininity when speaking languages that have grammatical classes associated with gender: "[Those languages] have endings that I have to change to feminine if I'm talking, and that does make me think about my gender quite a bit if I'm speaking those languages." She is aware that people's attitudes about language can vary widely, and uses the example of correctly pronouncing foreign words as something that can be perceived as cool and educated by some, but as "snooty" to others.

Quinn's favourite part of her body is her hair, which is very curly and long. It can be difficult to manage, she says, but if she puts in the effort she can do a lot with it. When she has just washed and styled it she can wear it out and it gives her a real boost to her self-esteem. In all her selfies, which are taken in more casual settings, she has it tied back loosely, escaping in random directions. Quinn's hair seems to correlate to her reported mood in each recording: The photo with the flyaways smoothed and tied down most severely (while at a university club meeting) is one where she felt "tired and sloppy", while those with more escaped ringlets and curls accompany happier recordings with friends.

Quinn has little experience with trans people or genders beyond the binary. ("I know there are people who are unspecified but I don't really know [about it].", and, when talking about her first experience with knowing a person transitioning in high school, "I found it slightly strange, it took me by surprise"). She is happy with being female herself, and associates femininity with positive feelings of pride ("[Femininity is] being proud of who I am as a women and showing that I am a woman in how I dress and my appearance.") She does not report much of a relationship with masculinity, and associates it primarily with the roles of men in society ("Especially in marriages, men have an important role as the head of the family and often the breadwinners.") Quinn thinks women have much more freedom and equality now when compared to her parents' generation, but it is not quite even, yet ("Society sees men as slightly more important than women even though our society has changed a lot.").

Reese - She/Her

In Reese's ItW recordings, she is professional and knowledgeable. In two of the three, she is in the role of teacher, guiding the person she is speaking to through some problem or assignment. In both cases, she asks open questions to the other person, asking them to explain some concept or process to her as a way to facilitate their learning. When it is her turn to guide, she speaks with authority. A bit slyly, she

[15] All three speak different Non-English languages.

talks about doing this even if she is not sure of the answer: “[I]f someone asks me something and I don’t know the answer I’ll just make up some reasonably plausible thing. It generally makes sense and is not always far wrong.” This confident, clear delivery comes through throughout the interview. I want to stress that Reese does not come across as egotistical or bull-headed – her answers to my questions always leave room for nuance, and she is excellent at separating her first impressions (often the social stereotype) from the reality of her experiences.

Reese’s selfies are top-down, capturing only a part of her face but most of her outfits, including shoes. In every image, she is wearing a black top and either khaki or black pants. She, like many other participants, talks about comfort being a major priority when she is picking clothes. Her relationship with her gender is described in a similar way -– comfortable. This comfort manifests in some apathy towards some of the gendered norms and expectations of fashion. She talks about using the men’s change room when it is closer, or having no problem checking the men’s section for shorts with pockets. “I suppose [my femininity’s] never been questioned so I’ve never really thought about it much.” She enjoys makeup but does not wear it most days because of the effort involved. She describes it as a “me thing”, rather than a "them thing", using it to give herself a boost of confidence when having to engage in public speaking.

Reese, like Mello, is athletic, and wears a necklace depicting her favourite sport. Her favourite body part is her legs, because of their muscular definition and strength. Reese identifies femininity with both physical appearance and with empathy. She talks about being cat-called, about how men “speak first and think later”, and how existing as a woman means she is “talked to with the expectation that you’ll always listen and always pander to someone.” Her impression of masculinity and maleness, in line with many other female and nonbinary participants, is framed in terms of this lack of empathy – masculinity is generally violent, grumpy, loud, and disconnected from family.

Reese is keenly aware of areas where Aotearoa New Zealand’s conception of gender falls short compared to elsewhere in the world, at least by her estimate. “New Zealand is behind the times” and “we’re [New Zealanders] quite conservative which is annoying.” But she is also aware of the colonial boundaries of some stereotypes, identifying short-shorts and skirts as “girl things”, while also pointing out that lots of people wear lavalava and that she would not necessarily use skirt-wearing as an indicator of femininity. She also frames gender in society as more about policing each other socially than inherent difference: it is not “women sit like this and men sit like this”, but women are told to “sit like a lady”.

Jesse - She/Her

> "[Makeup] makes you look a bit different and a bit more prim and proper."
>
> "I often wear dresses and tend to use my words and maybe my actions more, but if I was going out I would wear makeup and make more of an effort with my appearance."
>
> "I don’t like the fact that a lot of... I like modesty in clothing. I see this trend of females wearing less and less and less, and that’s trend that I don’t like to follow. There’s an expectation on females to dress like a barbie but I don’t feel it and guys have to be the muscly Ken."

Jesse's selfies follow a format: her face in the corner, usually sporting a huge grin, and the rest of the image taken up with her surroundings, the sky and various views of her flat (and flatmates). In half of them she is giving a thumbs up, in one she is closing her eyes against the sun, in another showing off a fuzzy bathrobe. They look like casual Facebook or Instagram posts, unpolished snapshots of her day. In all the photos where you can see her clothes, except the one with bathrobe, she is wearing a black hoodie. Like many other participants, she emphasises that she dresses for comfort more than anything else, especially at uni. When going back over the pictures during our interview, she mentions looking or being tired in most of them, though in the recordings she always sounds energetic and upbeat.

Many of Jesse's recordings are with her flatmates present, and you can tell she is very fond of them, embracing the social aspect of flatting that is such a huge part of university life in Ōtautahi Christchurch. She describes family not in terms of blood relations, but as "people who love you and want what's best for you" and you really get the feeling that her flatmates are considered and treated as her family. In one of Jesse's ItW recordings, the conversation briefly mentions "flat dads" and I bring this up during the interview as an interesting intersection of gendered expectations. When flatting, you have a bunch of young people together in a large house, often experiencing their first long-term stay away from home, and Jesse talks about how some people fit more naturally into certain roles. Flat dads and mums can be any gender, and you can be both a flat dad and flat mum for different things, though Jesse's context is flatting with other girls. Flat dads take on stereotypically masculine roles ("typical things that our dads would do [...] *dad sort of stuff*, like fix the car."): yard-work, fixing things that are broken, maybe contacting the landlord for larger issues. Flat mums, on the other hand, work inside and take on more caring roles, such as looking after the others after a night out partying.

Jesse's funny, recognising the humour in most situations and telling stories in a way that emphasises the silliest aspect. In one, she talks about going to get something out of her car in her bathrobe at the exact moment a bunch of people are walking by. In another, she recalls asking for directions to a rubbish bin at an airport in a foreign country, and being directed towards the international departures gate instead. For most questions, she has an anecdote or experience from her past to illustrate the point she is making. Her laughter is frequent and infectious. I get the feeling she is letting me in on the joke, laughing at a general situation or the reactions of the people involved, rather than at someone's expense.

Jesse does not really think about masculinity and femininity much at all, as it does not really affect her. She is happy being female, though there are aspects of the way society views and pressures women that she is uncomfortable with, such as the emphasis of sexualisation in fashion. She does not remember much explicit instruction or consideration about gender growing up, just that "In Year One boys were loud and a bit crazy ... in Year Seven boys were stinky!"

5.2.3 Binary Men

Noel - He/Him

> "Biologically [men and women] function correctly, and we balance each other really well."

> "We're all struggling to fit in. We all want to feel accepted and if we don't feel accepted we go home and think "Well, I'm not wearing that dress again" well *I'm* not saying that *[laughs]*, but you know what I mean."

> "Your opinions are only formed on the information that comes your way. If new information comes your way you might change your mind."

Noel is a man in his fifties, and forms the final part of the older triad with Alex and Kelly. He is incredibly friendly and has a bit of a salesman persona about him, which he mentions himself, as he works in sales. In his selfies, Noel's expression and position are similar across all but one – framing his head and shoulders, with his mouth slightly open in a small smile, looking down at the camera. There is one exception, a photo of the first recording taken by someone else: Noel is kitted up in the recording equipment, and is grinning hugely and a little self-consciously at the camera. Whether in his work uniform or in more casual clothes, his hair short and well-groomed, and always fully shaved, the overall impression from Noel's selfies is a friendly, clean-cut, self-assured professional. This presentation aligns with Noel's stated goals ("I try and dress a little bit younger, I'm conscious of that. I don't want to look old"). Noel's recordings all take place in his home, either on the phone or in person. In one, he is speaking to a flatmate about his personal history, and surviving a devastating event as a child. In another, he is trying to set up a friend with a woman he knows. In the last one, he is talking to one of his children about the goings-on of one of his flatmates.

In all honesty, Noel's interview made me uncomfortable in a way that none of the others do, through no fault or intention of his. Partially it was nervousness – Noel's was one of the earlier interviews, and the first binary male participant, and he presents a lot of the markers of the "kiwi bloke" stereotype referenced by a lot of other participants, both binary and nonbinary, when describing masculinity. At the time, I had not been living in Aotearoa long, and my experience with older New Zealanders was limited to my in-laws (in their 70s), what I picked up from the media, Pak 'n Save advertisements (a hugely exaggerated form of "kiwi bloke"-ness), and a handful of people within the university context, so I had very little experience with "actual" kiwi blokes. I think Noel made a lot of effort to make me feel comfortable and find common ground between us in a setting that was clearly uncomfortable for both of us, and I misinterpreted those overtures, and his more traditional views, as being a bit patronising. Listening back over the interview later, I can hear myself beginning to get flustered, and reasserting the boundaries of the semi-formal Question-and-Answer interview format (and, by extension, emphasising the researcher-participant relationship) in a way I do not with any other participants. I was distracted enough that I ended up forgetting to ask some of the questions.

Noel's children are grown, but he boards many students and younger adults and takes on somewhat of a parenting role with them. His approach to fatherhood is quite nurturing, and something where he explicitly tries to counter the masculine stereotypes of fathers as strict and distant rule-makers. His own father was quite competitive, and he describes his approach to parenthood as more deliberately nurturing. ("[Parents should] Let them develop. Don't snap their branches off."). Noel returns to this tree metaphor several times throughout the interview. Men are trees, according to Noel, women are the flowers that grow from those trees, and children are saplings trying to grow and thrive.

Taylor - He/Him

> "You see on the news reports, they just use the word "men" a lot, and "man", they don't specify who. Yeah, so I guess as a male I feel kind of attacked with the media."

> "My words are kind of jargon'd. Girls are normally pretty good at pronouncing words whereas guys sort of lazy words up."
>
> "To me, personally, I base [masculinity] around stereotypes. That's how I was brought up and that's how I want to be."

Taylor approached the research incredibly efficiently. He sent a schedule of when he intended to have the recordings completed by and stuck to it. During the interview, he responds to each question quickly and directly, and I can tell he has thought about the broad topics presented by the research materials quite a bit. It almost feels like he has prepared some of the answers despite not seeing the questions in advance, and the interview has the same smooth, efficient vibe I get from Taylor's emails. Though his voice and attitude are relaxed in many of the ItW recordings, his forethought and organisation comes through as well in their content – he coordinates a large group purchase, discusses the plans for an upcoming tattoo with a lot of elements, and goes over schoolwork with classmates.

Taylor has a relaxed and lighthearted relationship with his appearance and clothing. He lets his facial hair grow until he is tired of it ("it's like a time-lapse thing"). He likes to keep his hair short, but makes liberal use of hats when it starts growing out, to put off having to cut it so often. These hats, mostly snapbacks with sports team logos, are the exception to Taylor's laid back attitude to clothing; he enjoys team-branded sports gear ("it doesn't really go out of fashion ever because it just turns into vintage sports wear"). Taylor's playfulness in style comes in the form of deliberately "ugly" choices. In addition to a handful of deliberately ugly tops ("My ugly rugby jersey, that's what it's called. The idea of it is to look ugly."), Taylor is also sporting a deliberately terrible ("bogan") haircut he allowed some friends to give him part-way through making the ItW recordings. Specifically because of this haircut, he does mention being conscious of his appearance throughout the recordings, either thinking about getting the ugly haircut, or in being aware of how ugly it is after he has received it. I see both of Taylor's ugly rugby sweaters in his selfies, as well as the time-lapse beard, a few hats, a handful of t-shirts, and one white singlet (which he says he would only wear while alone and relaxing at home – even his playfulness around being a bogan has limits).

Taylor relates to masculinity deeply. Multiple times throughout the interview, when he invokes many of the kinds of stereotypes about men and women that exist, he is quick to qualify that he agrees with many of them ("[Masculinity is] like men do this, and men do that, which I'm fine with."). According to Taylor, many of these stereotypes exist *because* they have truth in them. He uses the example of blue being associated with masculinity: whether someone likes blue or not personally, *right now* it is a colour associated with males, and there is nothing inherently wrong with that association ("to me personally I base [masculinity] around stereotypes. That's how I was brought up and that's how I want to be."). Masculinity is a positive and guiding force in Taylor's life, and he expresses some frustration with the way the media and other forces tend to treat it negatively and without nuance ("femininity is all good and that, but the way they do it, they say femininity's good but masculinity's bad."). Growing up, Taylor said that messages about gender from most sources outside his own father were much less about male and female stereotypes, and much more about choice, but that there was always a feeling of some choices being better than others. Taylor rejects a lot of these ideas as he relates to his own sense of masculinity: "But to be honest, I did not really follow what they said [about gender]. They said [it's] like, choice-based, you can be whatever you want. But then they also made seem like it was negative to want to be like the man[ly] man."

Throughout the interview, Taylor's issues with the portrayal of masculinity and the messaging he received, and continues to receive, are all framed as they relate to him specifically and his own identity as a man. While it is obvious that he is frustrated with the lack of nuance in the media ("they're always blaming just men for this when it's only a certain amount of people"), I never get the impression that Taylor's angry at individuals or individual choices that don't align with his own. Since Taylor's frustration stems from the treatment of all men and masculinity as a whole as negative, it makes sense that he neither expects nor wants all men to act in ways that align with that stereotypical masculinity. When asked directly how he feels about people that do not fit into gender binaries, he is totally open ("it's their choice, it's all good with me"), and he acknowledges there are many kinds of ways to be a man, and that men who reject that stereotypical masculinity are still men.[16]

Riley - He/Him

> "Family is diverse. There's no normal in family."
>
> "I'm probably not self-conscious, but I'm self-aware."

Riley's the most soft-spoken of all the research participants. He often repeats my questions back to me when answering to make sure he understands. He is aware that trans people exist but has never met a trans person that he knows of. Riley is supportive of gay people and has many gay friends, a few of which are close enough that he considers them family. ("I feel fine with same sex... but I'm not sure about trans [...] I'm just not sure what to think of it. I'm neutral. I don't really know much about it so it's hard to judge it."). He take a measured, even response to most questions and his humour is self-reflexive, making a joke at his own expense before laughing.

When asked if he could change anything about his body, he says if it were a magical change he might make his stomach smaller, but does not really care much. ("If I wanted to change it enough I would probably go to the gym and change it but... my love for food? Or my love for being a slim handsome guy? *[laughs]*"). Riley's style is casual and sporty ("most of my clothes are like, ready to go to the gym, even though I don't go to the gym."), and he has a fairly laid back attitude to most things. His selfies reflect this, most of them showing off solid-coloured scoop-neck t-shirts or (unbuttoned) polo shirts, his hair unkempt but short enough that it looks intentional rather than messy. His photos are also missing the attention that some other participants have taken in regard to flattering lighting or angles. In about half of them, the entire top portion is blown out by indoor lights or the sun, and one of the others looks like it might have been taken a moment too late or early, his expression surprised, mouth partially open and not quite looking at the camera.

While talking about his relationship with someone in one recording, he says they get along partially because "we're both lazy and turn up late so we don't get upset at each other." He also is not bothered by harsh language or more offensive topics, though he does say he would tend to speak more gently, and would avoid making the same kinds of jokes with women (or gay friends) that he does with other straight

[16] Taylor's quote is in the context of vocal presentation, and I provide the full quote here as I think it may be open to different interpretations than the one I provide: "Most male guys probably sound like me, but sometimes other males, who don't like to identify as much as males, they typically have a different kind of voice. You get the assumption that they might not like "male", like "being male", like they're trying to be different to the stereotypical male, like they're trying to be their own kind of either male or like, transgender or something."

guys. He talks about how men tend to swear more often in his experience, and talk about memes or "rougher" topics: "I think men would talk more like insensitive topics, maybe more racist things or um maybe more, um, things that girls might think is wrong... homophobic, terrorism-related[17], or like [using the word] *retarded*, that kind of thing."

Like Quinn and Flynn, Riley is fluent in a language other than English and te reo Māori. He is acutely aware that his masculinity does not come across in this other language, or at least not in the same way as it does in English: "My manliness or whatever is not projected into [other language], and I feel like I'm speaking like a kindergarten kid or a little bit girly when I'm speaking [it]. I don't know why, and I don't know how to speak manly in [language]. I can try, but it feels a bit unnatural." This awareness of changes in the way he speaks is also apparent for Riley when he is speaking English in different contexts. With men, there is a bit of a need to impress ("maybe with girls or my gay friends I don't put on that kind of manly accent"), with his parents he speaks less overall, and with other bilingual friends he is aware of what he calls "chopping and changing" between languages. Riley reports deliberately avoiding this chopping and changing while wearing the recording equipment, as he assumed (correctly) that I was a monolingual English speaker.

Riley's attitude towards the perception of his masculinity also changes across languages. Friends that share his non-English language have pointed out "you sound so manly and cool when you speak English, but you sound so cute when you speak [other language]", but this does not bother him – Riley prefers to sound "natural" than to try to project a more masculine voice. Riley's aware of the differences in his voice both across languages and across contexts, and is much more sensitive to being emasculated or teased about these differences in English: "I was on the phone with a girlfriend and one of my male friends that was in the room was like "oh you sound like a little girl when you're talking on the phone who were you talking to?" and I was like "oh shut up!"".

Many of Riley's opinions throughout the interview reflect this balanced stance – an empathetic person who is aware that the ideas he grew up with might not always be correct, and that new experiences might change his opinion. He talks about growing up in a place without much diversity, with a lot of homophobia that he had to unlearn ("There's a homophobic vibe I would have grown up with that gay being like *"Ew!"*, kind of thing. But then I've learnt from other places, not from [some of my] family or whatever, that [being gay is] fine.") He speaks genuinely about his experiences with close gay friends, how unfair it is for them to have to worry about things as simple as holding hands in public. This recognition of his inexperience applies to his attitudes regarding his other language, as well: Riley thinks if he moved to a place where it was the dominant language, he might change the way he speaks as he develops more of a sense of gendered identity in that context. ("When I speak in English, there's the kiwi bloke kind of culture, and that's kind of in the background when I speak English, and I speak like a New Zealand male… But in [other language], there's no… none of that kind of kiwi bloke culture, so I don't feel the need to maybe like, be manly in [language]... if I moved there I might pick it up, I dunno.").

Riley's concept of his own masculinity is intrinsically linked to his heterosexuality. He describes himself as a male, and often clarifies that he is a not just male, but a straight male. He has never really thought about being male, as his masculinity is just an intrinsic part of his self image ("It's part of me, who I am, I can be a leader").

[17]This interview took place before the mosque attack in March, 2019. I do not believe Riley would talk about making those same kinds of jokes so cavalierly if this interview were after that date.

Likewise, he does not really think about his his physical appearance at all ("I think maybe I don't do a lot at all.. low maintenance… that's manly *laughs*"), except to avoid certain extra steps because they might make him look gay. Despite being aware of these markers to avoid them in his own presentation, Riley admits to not being very good at interpreting them in others, "[M]aybe, being a white straight male, I assume everyone is straight from the beginning. I don't pick up if someone is gay." I get the impression that Riley's concern with being perceived as gay is less because it is a pejorative association for him, but more because of its inaccuracy, as his heterosexuality and masculinity are so intertwined.

Jordon - He/Him

> "[I was taught that] Boys do sport and work in construction whereas girls are nurses and don't play so much physical sport like rugby."
>
> "Particularly when I was going through puberty I was a little like… having a high-pitched voice was difficult, and I would step my voice down sometimes. I don't know whether or not that's become like a habit and changed what my voice is like now, or if this is what my voice would naturally be, but I don't make that conscious decision anymore."

Jordon's interview is one of the shortest, and plagued by technical issues. It is interrupted multiple times by needing to change the batteries or rerecord some portion. This may have contributed to the general atmosphere of the interview as being slightly cooler, and never quite losing the stiff, formal edges between each question and answer that happens in the other interviews. Jordon's responses are generally quite short, and there is a hesitation in some of his responses, like he is leaving something unsaid. He answers many of the questions with a "yes, but I can't really explain how", or a "yeah nah, I'm not sure." At the time, I remember wondering if he was impatient to leave or unhappy with the way the interview was going. However, in revisiting his recordings and looking at the context in hindsight – the first few weeks of his first semester at university – I think it is more likely he was either nervous, or a little overwhelmed by the number of steps and recordings involved in the research, or both. I emphasised at the start of the interview and reiterated throughout that there were no "correct" answers to the questions. Jordon, among the binary participants, was the most frequent to say they did not know or could not answer a question, without providing more of a reason. I wonder if this was to avoid saying something he thought I might interpret as "wrong" because it went against the narratives in popular culture, as pointed out by Kelly and Taylor explicitly, that conflates toxic masculinity with masculinity in general.

Jordon's relationship with masculinity is incredibly strong, and he speaks highly of influential role models in his life teaching him about positive masculinity. These role models were careful in separating positive forms of masculinity from the more toxic and negative associations of patriarchy: "[Masculinity] isn't about asserting dominance, but it's about being responsible, and making sure that anyone who depends on you, *can* depend on you." His recordings are a very typical slice-of-life of a first-year undergraduate student: chatting with friends, discussing movies and future plans while watching Netflix or over drinks that he (suspiciously) insists to the recording equipment are nonalcoholic. His selfies are similarly generic: trainers, gym socks, nice khaki shorts and collared shirts or o-neck tees. While Jordon is not aware of any clothing choices he makes specifically to present his masculinity, after some

thought he admits to avoiding shirts with low-cut or plunging necklines, though this might be more about avoiding certain presentations of masculinity.

Though not something he dresses for specifically, a masculine presentation is an important part of Jordon's relationship with his body. He goes to the gym regularly to achieve this ("Part of my influence for exercising was to do with my masculinity, and for feeling more like an adult man"). When asked about his favourite part of his body, he says it is his chest, because it looks masculine and fit (he jokes "I like to look at myself because I'm vain.") When asked if he does anything to show off his chest, though, Jordon says no, he tries to avoid it, particularly because he has seen other men do it, and it comes across as "transparent." This idea of pursuing a masculine presentation, and being proud of that presentation, without being overly boastful, is important to Jordon.

Devon - He/Him

> "To be feminine you are caring and kind maybe a bit dainty. You can do things but when it comes to labour or something you may need a bit of help. [...] It's just the idea of like "girls don't poop" – they hold themselves to a pristine standard to be feminine"
>
> "Being a man, I'm not wanted as much."

Devon is from rural Aotearoa New Zealand, and is attending university in a highly competitive STEM field to pursue a childhood dream. He gets visibly and audibly tongue-tied when talking about gender and relationships. When I asked whether his feelings about his gender have changed over time, Devon is the only participant to frame it as a practical matter of pros vs cons. As a kid, he reports, Devon did not worry about gender, but as he got older he realised there were other options ("I thought it might be like, alright to be a girl"). He explains that the benefits of manhood often come later in life, in better jobs, and having more authority and power at work. When you're younger, though, it is easier to be a woman ("As a guy you're always constantly worried about sizing people up or social status or like size [...] As a girl, I could get more scholarships and get into bars."). Ultimately, though, Devon came to the conclusion that being a man is the correct choice for him ("I'm more happier being a guy than I would be as a girl."). Masculinity for Devon is also associated with loneliness and a lack of desirability, especially as a young person ("being a man, I'm not wanted as much").

Devon is a scientist, through and through. In one recording he makes fun of an ex for being into astrology, and makes fun of himself for trying to make it work with such a person. When asked about gender and sex, he takes a similar stance, saying that you can be whatever you want to identify as, but biology is not something you can change really, scientifically. The word he uses to refer to women most often is "females". Devon wears braces and reading glasses, and approaches them both with the same scientific practicality, rather than as a means of self-expression or fashion. He has never treated them as an accessory or customised the colours ("they're there to serve a purpose"). With clothing, he tries to match colours but does not consider much beyond that.

Physical presentation outside of fashion, though, is important to Devon, and he takes care with personal grooming. In discussing the context of every ItW recording, he talks about being aware of whether he needed to take a shower, get a haircut, shave, etc. he is self-conscious about his acne, wears a rash-guard swimming, and

tends to avoid taking his shirt off even inside his own apartment. He deliberately deepens his voice around other men to appear more masculine, and is conscious of his pitch rising when he is excited, and tries to keep it in check because "it doesn't sound the greatest."

This is painting a picture of Devon as a stereotypical "shy guy into science" that fits, but is not his whole story. Devon is passionate about space and loves to go to the observatory, he loves his curly hair and his eyes because their colour matches his real name. He seeks out new experiences and friendships, and, at least at the time of our interview, was embracing university as a way to experience new things and meet new people. His self-consciousness reflects a keen attention to detail, and he was one of the only participants to discuss how his interlocutors were feeling, as well as his own emotional state, when asked about the general tone of the ItW recordings.

5.2.4 Conclusion

The purpose of this chapter was to give a brief introduction to the participants involved, to represent them as more than simply p-values or the points on a plot that appear throughout the following chapters. There is so much to each of them that I have left out: Quinn's deep relationship with music, Ollie's love of his hands, Taylor's speculation about the MCU,[18] Kit's craftsmanship with costume design, Kelly's story about dressing "a little too butch". During Alex's section, I mentioned that the Alex portrayed through the ItW recordings and interview is just a snapshot, a small slice of the Alex I have come to know over the last few years. In the same way, a handful of paragraphs can never express the depth and nuance of a whole human, or even the slice of human presented in a few hours of audio recordings. I have tried to let each participant's words stand on their own, and to be clear where I have filled in some blanks with my own interpretations.

There is so much, too, that I have left out because it is too specific to identifying participants, or impossible to present with any objectivity: The heartbreaking moment(s) where a participant discloses personal trauma, or says something bigoted. Above all, each of my participants is a person worthy of respect. While I must be honest and avoid presenting them in a misleading light, it is also not my place as a researcher to disparage these participants or present them negatively. Consequently, I have erred on the side of "If you can't say something nice, don't say anything at all", both in my own interpretation and in presenting their own words.

Some broader trends became apparent to me over the course of the interviews, that did not occur to me beforehand. The effects that "pro-woman" forms of feminism have on young men, for example, is something that many of the binary men (and of the older participants with male children) comment on. Many expressed feeling like the positive aspect uplifting women and girls is coming, unnecessarily, at the expense of men. In a similar way, many of the men report feeling disconnected from femininity, or like they have no relationship to it (of all the nonbinary and binary female participants, only Quinn reports having no relationship to masculinity). Something else that did not occur to me as a relatively able-bodied person, is the way (dis)ability and pain have a major effect on people's relationship to their bodies (and, by extension, gender). I have never really had to consider the effect of clothing or the environment/weather on my body in the same way that many of the participants of all gender groups must. Making choices about the kinds of shoes or jackets to

[18] Marvel Cinematic Universe

wear, or whether to bind or put your hair up or not, takes on an extra complexity and difficulty when the result of that choice may be days of pain or discomfort afterword.

I have taken my time in contextualising the quantitative and qualitative analyses that follow. The detail involved in presenting the participants and research context is intended to complement the following chapters, in which these sixteen humans are often represented as a collection of dots and numbers, the output of statistical outcomes and digestible visualisations. This approach results in the same balancing tension I identified throughout the literature in Chapter 2, between the physiological and social, micro and macro, measurement and intention. Neither is completely separable from the other, and attention to both is critical in fully understanding the complex ways these participants navigate the social foundation and their individual context in presenting their gender and broader identity to the world.

Chapter 6

Pitch

6.1 Introduction

This chapter addresses differences in pitch and pitch range between nonbinary and binary speakers. Due to pitch's salience as a gendered marker, I begin with a selection of participant quotes regarding their use of pitch in Section 6.2. In Section 6.3 I present the relevant hypotheses and predictions. In Section 6.4, I describe the methods by which the data was prepared for analysis and subsequently analyzed using two different tools, Praat and Talkin (2015)'s **R**obust **E**poch **A**nd **P**itch **E**stimato**R**. (REAPER). I include a comparison of both tools and my motivations for using both in subsequent analyses. Section 6.5 presents the results of various analysis on the pitch of speakers, including summary statistics, an exploration of modes across speaker groups, pitch distributions, cross-contextual variations, and some qualitative examples of pitch manipulation and control by nonbinary speakers. Finally, in section 6.6, I discuss the results and return to each hypothesis to address how the results align with the predictions.

6.2 In Their Own Words: Pitch

In the review of the literature in Section 2.2.3, I presented experimental evidence showing the link between the perception of sex/gender and speaker pitch is robust, much more so than the link between speaker physiology and pitch (in particular, overt markers associated with sex/gender like overall body size or presence of facial hair are much weaker predictors of pitch than more subtle, less visually obvious physiological factors like pharynx length or testosterone production[1]). Before addressing the hypotheses, I present here a handful of quotes from the participant interviews that emphasise this link is directly salient to the speakers in this study. While non-phonetic linguistic markers, such as the use of profanity, topic differences, and the overall amount of speech were frequently identified by participants as markers of gendered difference in speech, pitch was by far the most widely identified acoustic marker.

The purpose of this small introductory section is two-fold. Firstly, to establish that pitch is both highly salient, and highly gendered, to both speakers and listeners. Therefore, it is eminently available for use in presenting incongruence. Secondly, to spotlight some of the ways that participants in this study, and nonbinary participants in particular, are aware of the way pitch (and by extension, the pitch of their own voice) is gendered perceptually by the people they are speaking to. This awareness leads to some incredible insights into their own ability to manipulate this aspect of the speech signal. As in other areas of this thesis, I have done some some minor

[1]Though the growth of the larynx during puberty can result in a prominent Adam's Apple, it is visually less prominent for larger speakers.

editing of the quotes, such as removing some hesitations and speech errors, with the goal of maximising clarity. I present the full transcription in a footnote where it differs substantially from the presented quote.

> "Masculinity is just your classic stereotypes: Deep voice. Tough talk. Talking smack. [...] [Girls have] a higher pitch and all that."
>
> — Taylor (Binary Male), Interview, 54:38[2].

> *J: Do you think people can tell that you're nonbinary?*
>
> "I think in terms of the way I talk, anyway, unless I'm being really flamboyant, my speech is actually quite deep."
>
> — Avery, Interview, 57:40.

> "Pitch alters sometimes like, if I'm with a group of men I'll, you know, *[lowers pitch]* speak more kind of like this. "
>
> — Istus (Nonbinary), Interview, 16:03.

> "I sound really female because like, I'm always surprised by my voice when I hear it on a recording. It's like, that's not my voice it's too high-pitched and girly sounds and yeah."
>
> — Alex(Nonbinary), Interview, 25:05.

> "If [person]'s[3] having a masculine day and is going by he/him [...] I think that [person] subconsciously deepens [the] voice a little bit, and I feel like I might've subconsciously done that as well but it's not like I [said] *I'm nonbinary now I'd better change my voice up!"*
>
> — Kit (Nonbinary), Interview, 1:23:28.[4]

> *J: How do you use language when presenting your gender, or identifying gender in other people?*
>
> "The first thing I'd think of that I definitely [do] would be lowering my pitch to be taken more seriously which sounds really sexist but, you know [...] A woman with a lower-pitched voice or a man with a higher-pitched voice I'd be more questioning their gender or their sexuality, yeah."
>
> — Alex (Nonbinary), Interview, 23:54.[5]

[2]Taylor: Um so yeah, masculinity is just your classic stereotypes like, um, deep voice, tough talk, talking smack all that kind of stuff. Um yeah just kind of easy, chill, um yeah with the with the roast and all that." (*J: And what does femininity mean to you?*) Taylor: Yeah its about the same so like, just how, like, you would think as your average stereotypical girl but not like the um, the classic ones you see in the movies probably, just more like chill, talkative, like they talk about different things and like, they do talk about more like, girly kind of stuff, like clothing or um, yeah, that kind of stuff, and higher pitch and all that, yeah."

[3]This refers to a genderfluid friend of Kit's. I have obfuscated Kit's use of this person's pronouns in the excerpt to protect the person's privacy.

[4]"If [pronoun]'s having a masculine day and is going by he/him like [pronoun], or [pronoun], sounds a bit like, I think that [person] subconsciously deepens [pronoun]'s voice a little bit, and, but, like, and I mean I feel like that I might've subconsciously done that as well but it's not like I, it, it wasn't an *ooooh I'm nonbinary now I'd better change my voice up!"*

[5]*(J: How do you use language when presenting your gender, or navigating gender would be another way to frame it, when identifying gender in other people?)* Alex: Yeah um, I mean, I guess most of it is just to do with pitch and things like that, is be, sort of, like the first thing I'd think of that I definitely, yeah, would be lowering my pitch to be taken more seriously which *[both laugh]* sounds really sexist, but you know... and, yeah, I guess I'd also like, probably be making assumptions about people's genders just based on the pitch of their voice a bit like, yeah, a woman with lower-pitched voice or a man with higher-pitched voice is sort of I'd be more questioning about their gender or their sexuality or... yeah."

> "When I first met [person][6] I tried to keep my voice very, I don't know, low or whatever, but um... Sometimes if I'm talking to [person] I'll just slip and I'll be like *shit shit shut up shut up*, you know, *don't talk anymore!* I get really self-conscious around [person] but, um, around other people, not really."
>
> — Ollie (Nonbinary), Interview, 13:57.

6.3 Hypotheses

This section outlines the predictions made regarding the utilisation of gendered incongruence in pitch. Since pitch is characterised as a gradient signal, Hypothesis **H1b** (3.3.2) applies. The measures of interest are aggregated over time and context via pitch means, and absolute difference across peaks and troughs via pitch ranges.

6.3.1 Hypothesis 1b: Gradient Signal Incongruence

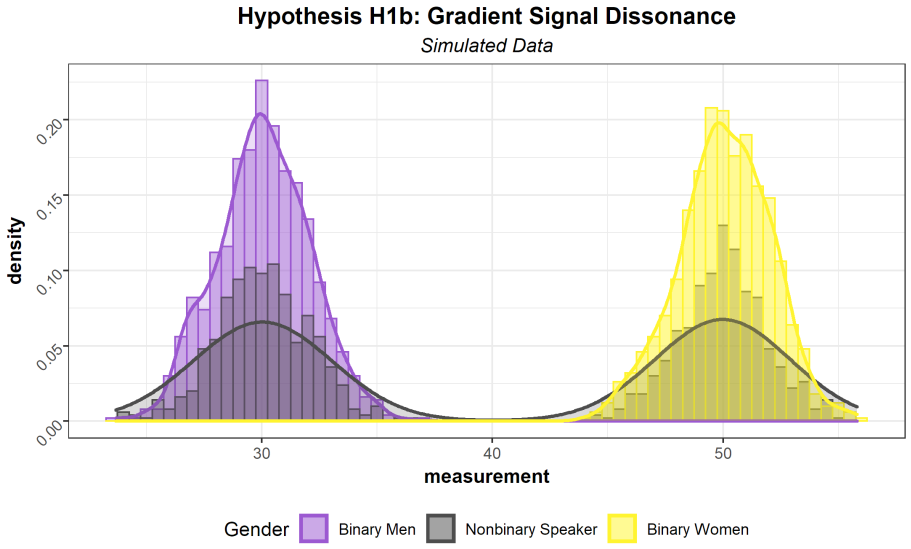
Given a single gradient variable, nonbinary speakers can communicate identity by producing incongruence either via **ambiguity** *in the form of more centrally-distributed tokens compared to binary male and binary female speakers, or via* **dissonance** *by producing tokens in a bimodal fashion, where a single speaker's productions are within both binary male and binary female ranges.*

I reproduce Figures 3.2 and 3.3 from Section 3.3.2 here in Figure 6.1. **H1b** predicts that nonbinary speakers producing incongruence may do so in one of two ways. Incongruence via dissonance (Figure 6.1i) is achieved by producing a pitch range whose two peaks are each within the expected values for men and women. Incongruence via ambiguity (Figure 6.1ii) is achieved by producing a pitch range with a single peak located between that of men and women. Both forms of incongruence predict that a nonbinary speaker's mean pitch should fall between that of men and women, but for different reasons: Incongruence via ambiguity achieves a centralised mean via centralised tokens with a small range, while incongruence via dissonance achieves this mean as a product of the tokens being arranged bimodally with an overall wide range due to the space between peaks. I present these predictions using values extrapolated from measurements in the literature in Table 6.1, though the critique I made in 2 of these values, particularly for binary women being extrapolated from small numbers of samples, is still relevant.

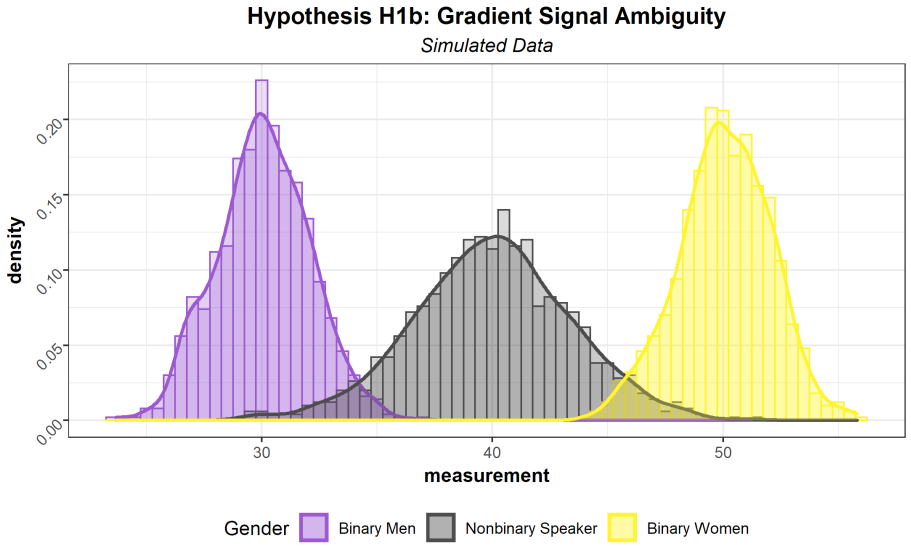
Speaker	Predicted Mean F0	Predicted Mode(s)	Range (Hz)	Range Width (Hz)
Incongruence via Ambiguity	∼160Hz	1	120 – 200	80
Incongruence via Dissonance	∼160Hz	2	90.5 – 256	165.5
Masculine Norms	∼120Hz*	1	90.5 – 165.2*	74.7
Feminine Norms	∼200Hz*	1	179 – 256*	77

TABLE 6.1: Values for binary gender norms, marked with a *, are taken from Baken and Orlikoff (2000, pp. 175–176), indicating spontaneous speech of men and women between 17.9 and 26 years old speaking American English. Values for nonbinary speakers are predictions extrapolated from these norms based on hypothesised incongruence.

[6] At the time of the interview, this is a person Ollie met relatively recently, that Ollie is not out to.



(i) Predicted Pitch Dissonance



(ii) Predicted Pitch Ambiguity

FIGURE 6.1: A reproduction of figures Figures 3.2 and 3.3 from Chapter 3, illustrating predicted gradient signal incongruence. While the reproduced image is generalised for all gradient signals, when addressing pitch specifically the X-axis here can be read as a measurement of Hz and the Y-axis as a measurement of density.

6.3.2 Hypothesis 2: Variation and the Immediate Context

Immediate context has a greater influence on nonbinary speakers than binary speakers.

H2 predicts that nonbinary speakers will exhibit greater variation than binary speakers across contexts. In this chapter, for reasons I explain in Section 6.5, I limit the scope of "context" here to refer to different recording contexts.

6.3.3 Hypothesis 3: Variation and the Individual Condition

Variation is bound by the individual circumstances of each participant. Individuals with shared socially-encoded circumstances, such as age and binary gender, will speak more like

each other.

H3 tempers the predictions made by **H1** in a number of ways. I highlight two here: "natural" mean pitch, and preexisting social factors.

Physiology sets limits on pitch and pitch range in a number of ways: First, it sets absolute boundaries in the form of limiting the sounds that are possible for a given body to produce. However, as established in the review of the literature in Section 2.2.3, the link between pitch and physiology is not absolute, nor is the division between the physiological and the socially malleable. In fact, social and other contextual factors (such as the language being spoken) have been shown to exert a substantial influence on pitch production (Ng, Hsueh, and Sam Leung, 2010; Pépiot, 2015). More critically for this project, physiology in combination with other aspects of the individual context sets softer boundaries in terms of what is comfortable for speakers to produce, as well as what is considered possible or probable both for speakers to produce, and listeners to perceive as "natural". I use "natural" here to reference an ease, regularity, and comfort in production. One of the participants in this study, Istus, contrasts her "natural" voice with her "normal" voice, and exhibits her ability to seamlessly switch between them. Neither of these voices sounds abnormal or unnatural, but they exhibit different characteristics. The "natural" voice is part of the individual condition, and serves as the canvas on to which other vocal changes may be made. These softer boundaries established by the individual condition set limits on what can be produced without undue strain or difficulty on the speaker's part, while also not being so far from modal speech as to register perceptually as another phonation type (such as falsetto or creak) to most listeners. [7]

H3 also predicts that pitch and pitch range will be tempered by other factors, including social identity and participant history. While this is predicted to hold true for all participants regardless of gender, it specifically predicts that nonbinary speakers will vary more within-group than binary speakers, due to both the wider variety of identities encompassed in the nonbinary gender category, and to the assumed variety of difference in personal experience and history of the nonbinary participants prior to their participation in this research.

6.4 Methodology: Praat and REAPER

6.4.1 One Size Fits None

This dataset contained a number of inherent issues that, while not unique in isolation, combine to make some forms of established analysis difficult. While examining overall pitch has the benefit of providing a "bigger picture" snapshot of the voice without the need of transcription or phonemic alignment, it is also difficult to isolate speech from other sources of noise. The methodology used to obtain the In the Wild (ItW) recordings was suitable for addressing both privacy concerns and obtaining clean single-speaker recordings, but it was not infallible: a handful of recordings made in a vehicle, for example, contain low-frequency engine noise, changes in speech consistent with the Lombard effect (Brumm and Zollinger, 2011, p. 1107), or (in two cases) spikes of unpredictable high-frequency noise associated with airbrakes. Recordings where the main participant is speaking to interlocutors that are very old, very young, or nonhuman contain obvious shifts in the participant's register, such

[7] I use terminology regarding phonation broadly here, with modal referring to "the range of fundamental frequencies that are normally used in speaking and singing" (Hollien, 1974, p. 126) and creaky referring to a number of nonmodal phonation types that often (but not always) share 2/3 of the acoustic properties of low F0, irregular F0, and glottal construction Keating, Garellek, and Kreiman, 2015.

as increased amplitude, reduced speech rate, or wider intonational curves. The very variety that makes these natural[8] settings appealing for study are their own drawback if the goal is a comparative analysis, as mine is here.

In some recordings, there are very obvious contextual explanations for variations in pitch measurements, such as lots of overlapped speech with children (This can be seen clearly in Recording #4 in Figure 6.12, in which the participant, Alex, is talking to two young children). In other recordings, there are differences that appear to align with the predictions made in **H2**, in being driven by contextual factors. However, it is often impossible to determine *which* of multiple possible factors is the real source of difference, or if the variation only exists as a consequence of multiple influences, such as a participant whose recordings with family overlap exactly/exclusively with their recordings in more public settings. I have highlighted what I believe are the relevant contexts when examining recordings in detail, and provide a general overview of each participant's recording context in Appendix B. I have not, however, explicitly laid out the context and possible sources of contextual variation for every recording for every speaker, both in consideration of participant privacy and because it is an impossible task to objectively and reliably account for every possible contextual clue.

Some forms of analysis, including linear mixed effects models and generalised additive models, were considered but determined to be unsuitable for this data for a number of reasons. Primarily among these was the wide variation across and between ItW recordings, and the non-uniformity of the pitch distributions across participants. While I did make some attempt to clean the data to address some of these issues, such as normalising via semitone conversion and scaling, ultimately I determined the diversity of the data made reshaping it to suit the parameters required of those statistical analyses inappropriate.

6.4.2 Pitch Evaluation and Measurement with Praat and REAPER

This thesis uses two ways of extracting the pitch measurements: Praat and REAPER. Praat (Boersma and Weenink, 2016) is the most commonly used phonetics software for tracking F0 (Strömbergsson, 2016, p. 526) and many other forms of phonetic analysis. The other pitch measuring tool used was REAPER, a newer, command-line only program designed to measure pitch, voicing, and glottal closures. In this seciton, I first explain some of strengths and weaknesses of the established program (Praat), and what led me to consider an alternative system of extracting pitch measurements. I then briefly compare the differences in the way the two programs measure pitch on a functional level. In section 6.4.2, I outline the specific process I used to extract pitch using each program, followed by a section comparing the two methods and the criteria used for choosing the most appropriate one at each stage of the analysis. Ultimately, based on these comparisons, I conclude that both provide suitable data for different types of analysis, and explain where I have chosen to use each in addressing the research questions and hypotheses.

Process Differences Between Praat and REAPER

To measure pitch, Praat uses an autocorrelation algorithm, outlined in Boersma et al. (1993), that samples each time window four times to determine the most likely pitch period. Values below the pitch floors and above the pitch ceilings are discarded before sampling. The silence threshold (the amplitude cut-off for determining if a

[8] As natural as one can be when wearing obvious equipment and the knowledge that the conversation is being recorded.

token should be considered voiced), the voicing threshold (measured by the strength of the unvoiced candidate relative to the maximum possible autocorrelation, from 0-1), the pitch floor (in Hz), and the pitch ceiling (also in Hz) can all be set by the user, as well as a number of settings related to though the program provides defaults as well. The window size cannot be set directly, but is determined by dividing the pitch floor by 3. Errors in the pitch measurement can lead to pitch periods being collapsed (resulting in pitch halving) or split (resulting in pitch doubling). These errors are often caused by inappropriate floor/ceiling settings and window sizes, but can also be the result of the algorithm inappropriately dealing with other phonation types, such as creaky voice (Styler, 2013). Praat does not have a good, automatic way of reducing these errors, or, crucially, identifying them. This makes it difficult for a researcher to determine whether measured changes in pitch are caused by tracking errors or legitimate changes in speech without spot-checking, so reducing the number of errors usually involved individually examining and tinkering with the parameters for each speaker. Despite this, Praat also benefits somewhat from the sheer ubiquity of these issues: It is relatively easy to address pitch doubling/halving when you know what to look for, and individual window sizes and pitch thresholds can be easily customised for a given recording to maximise results.

REAPER also uses autocorrelation, via a dynamic programming approach. It first demarcates timepoints every 0.005s, and within those timepoints identifies glottal closures via finding negative amplitude peaks in the signal, labelled glottal closure instants (GCI). The program than creates what Talkin calls a lattice structure between consecutive GCIs and the possible minimum/maximum pitch values occurring between them (Talkin, 2015). Beginning with the final pulse in the file, REAPER works backwards, calculating the likelihood at each timepoint of whether the signal is voiced or unvoiced. Within the voiced possibility, the likelihood of each possible pitch between the min/max at each time point between GCIs is determined, and the pitch value with the highest likelihood is output.

To compare directly (though the methods are quite different, and this explanation simplifies what is happening in both programs): Praat's autocorrelation samples each window four times, regardless of the window size (which is set by the end user), REAPER, on the other hand, fixes pitch measurements at 0.005s, but the effective window size (over which correlations are computed) is dynamic, centred on the GCI pulse of each period.[9] Pitch tracking errors are effectively erased in this method, since huge jumps in pitch across such small time measurements would be given incredibly low likelihoods.

REAPER provides a much richer and more reliable output, especially in the lower frequencies, including glottal pulse information, and avoids many of Praat's doubling/halving errors. However, this depth of information comes at a huge computational cost, which increases exponentially given the size of a recording (since it must identify *all* pulses before measuring pitch). REAPER also lacks Praat's level of flexibility to easily or manually alter settings, due to both the processing time and its lack of a GUI (graphical user interface), as it is run via the command line. It is also less suitable for field recordings, like the ItW recordings in this thesis, or those that have undergone filtering (such as noise removal) as it is designed for use with studio-quality speech. Finally, its way of estimating the correct pitch via dynamic programming and backtracking makes it very difficult to ascertain exactly how each

[9] A full explanation of the processes and complex mathematical functions involved in a complete comparison of Praat and REAPER in pitch extraction and detection of glottal pulses is far beyond the scope of this thesis, and in fact is a thesis in its own right – See Dorreen, 2017 for a more detailed analysis of the bones behind REAPER.

determination was made at each time point, and REAPER's high time resolution means that identical analyses run on different hardware results in slightly different outputs for each timepoint.

Pitch Extraction using Praat

Pitch measures were extracted using Praat, via Fromont and Hay (2012)'s Language, Brain & Behaviour Corpus Analysis Tool (LaBB-CAT) web interface, with custom pitch settings. Custom settings were used because the defaults for measuring pitch employed by Praat and LaBB-CAT differ slightly: Praat's default/standard settings set the pitch floor at 75Hz and the pitch ceiling at 500Hz for all recordings. LaBB-CAT's default settings at the time of analysis first check to see whether the input contains something that looks like a "sex/gender" column[10], and, if interpretable, applies a default for men's voices of a 30Hz floor and 250Hz ceiling, and a default floor of 60Hz and a ceiling of 500Hz for women's voices. If no gender or manual value is specified or able to be interpreted, LaBB-CAT uses the female settings[11]. Rather than employing these defaults, I set a pitch floor of 30Hz and a pitch ceiling of 500Hz, encompassing both of LaBB-CAT's default settings for binary gender. Using LaBB-CAT's in-built Praat functionality and defaults, each utterance by each participant was split into 1-second chunks, and summary statistics (minimum possible pitch in Hz, maximum possible pitch in Hz, and mean pitch in Hz) for each chunk was extracted. The voicing threshold was set at 0.45 and the window offset was 0.0s. Praat's defaults were used for all other settings.

Using such a wide pitch range and narrow window increases the likelihood of doubling and halving errors, but my justification for doing so was to more equally share those error rates across participants, rather than reduce errors for one group at the expense of the other two. Using one set of binary defaults for all groups begs the question of "which measures do you treat as default?" and would therefore create unbalanced error rates. I also considered setting three separate group limits for men, women, and nonbinary participants, as well as determining individual limits for each speaker, but decided against these approaches for a number of reasons: From a practical standpoint, there is a huge loss of efficiency and increased chance of error involved in using 16 individual sets of parameter settings, and the need to change these 16 setting arrays for every subsequent analysis.[12] The lack of precedence in the literature for measuring nonbinary voices, or even nonstandard binary voices, and the preponderance of using defaults (whether a single measure across all participants or binary gender settings[13]) make it difficult to know the likelihood that any given setting array would inadvertently exacerbate tracking errors. Finally, while the primary goal was determining the best way to measure pitch for these particular

[10] Praat analysis uses .wav files of recorded speech directly, while LaBB-CAT reads spreadsheet (.xls) files that contain meta information as well as a link to the location of a given sound within the corpus.

[11] While this default behavior has not changed, LaBB-CAT has since updated its interface to make this application of defaults much more transparent and allow users to easily specify new columns and values.

[12] I also considered a mixed approach such as using LaBB-CAT's male and female defaults while setting bespoke values for nonbinary participants. This seemed like a "worst of both worlds" approach, in that it would reduce the rate of tracking errors across the board, but unevenly (presumably reducing errors in nonbinary speakers but increasing those in the more "nonstandard" binary voices), would still be less efficient than a single measure, and still introduce/increase the possibility of researcher error in future measuring. This latter point would apply unevenly, as only nonbinary speakers would require the settings array be customised for each analysis.

[13] Vogel et al. (2009) found that using generic one-size-fits-all or binary sex-specific settings tended to be equivalent to bespoke per-speaker settings when applied to large-scale analysis (>1k files).

16 speakers, the bigger-picture objective of determining which tool may be best for incorporating the analysis of nonbinary voices into existing methodologies meant I wanted to compare them "out of the box", and make it easier for future additions to the RAINBO corpus (or other researchers doing similar work) to replicate these analyses, whether they use Praat, LaBB-CAT, REAPER, or something else.

Pitch Extraction using REAPER

While LaBB-CAT has REAPER functionality built in, it works best for projects containing relatively clean, continuous speech from a single speaker (such as narratives or reading passages). It is inappropriate for use in RAINBO, particularly the ItW recordings, due to the presence of multiple speakers, long silences, and environmental noises. I therefore prepped these recordings and ran the REAPER tool outside of LaBB-CAT.

Due to the increased processing times for larger and longer files, preparing recordings for REAPER required some extra steps to reduce the file size. Speech that was not already transcribed, like the interviews with binary participants, was annotated for each speaker utterance in ELAN (The Language Archive, 2021), these timestamps were then exported to textgrid files. Each recording was converted to mono and downsampled to 16Hz in Praat, with participant utterances reconcatenated into a new, smaller wav file. With the REAPER program, via the command line, pitch measurements were extracted from the wav files every 0.005 seconds and exported to a text file. From REAPER's output, periods of glottal closure and pitch values above 300Hz were removed, leaving 7,365,432 data points across all participants and recording types, roughly split between ItW and interview recordings (3,766,651 data points across interview and 3,598,781 across ItW recordings).

6.4.3 Comparison of Praat and REAPER

Table 6.2 shows the summary statistics of measurements obtained from REAPER and Praat[14]. The final two columns, Pitch Range and Mean F0, show the difference between the observations from both programs, arbitrarily taking the REAPER values as the starting point. Pitch ranges inside $\leftarrow \rightarrow$ arrows represent a pitch range that is expanded in the REAPER measurements by that amount, while ranges inside $\rightarrow \leftarrow$ arrows are contracted by that value. These arrows were chosen to better visually represent how pitch ranges widened or narrowed between methods of measurement. In the Mean F0 columns, values listed beside $\downarrow$ indicate how much lower REAPER's estimate of mean pitch is compared to Praat, while the value beside the $\uparrow$ indicates that REAPER estimates the speaker's pitch to be higher than Praat's measurement. For example, Remi's pitch range difference is listed as $\leftarrow$ 128.26 $\rightarrow$ and her mean F0 difference is listed as $\downarrow$ 36.8, indicating that the values derived from REAPER's pitch measurements estimate Remi's pitch range to be 128.26Hz wider, and her mean F0 to be 36.8Hz lower, compared to the values derived from Praat.

For every speaker except Riley, REAPER's data shows a mean pitch in the range of 1.76-39.85Hz lower than Praat's estimate. When looking at pitch ranges, measurements made with REAPER tend to be larger, though there are a few exceptions, as five speakers (two men and three nonbinary speakers) show a contracted pitch range

[14]The extraction and analysis of the REAPER data was interrupted by the COVID-19 lockdowns and had to be re-run on a different computer. Due to the stochastic nature of REAPER's measurements, explained in the previous section, the data reported in this Section varies slightly, in the realm of a few Hz, from the data reported elsewhere in the thesis.

		REAPER						Praat						Difference (REAPER - Praat)	
Participant	Gender	Obs	Min	Max	M.A.D.	Mean	Range	Obs	Min	Max	M.A.D.	Mean	Range	Pitch Range	Mean F0
		All values except number of observations (Obs) listed in Hz													
	Women														
Remi		533 401	0.45	351.11	70.13	175.78	350.65	3558	101.38	323.77	44.48	212.58	222.39	← 128.26 →	↓ 36.8
Kelly		595 230	53.98	271.59	43.52	162.78	217.61	3490	94.16	294.31	40.03	194.24	200.15	← 17.46 →	↓ 31.45
Quinn		232 576	78.66	257.87	35.84	168.26	179.22	1404	125.73	273.99	29.65	199.86	148.26	← 30.96 →	↓ 31.6
Reese		274 973	10.59	325.83	63.05	168.21	315.24	1615	127.61	286.99	31.88	207.30	159.38	← 155.86 →	↓ 39.09
Jesse		478 563	102.92	276.71	34.76	189.81	173.78	2858	160.58	294.01	26.69	227.29	133.43	← 40.35 →	↓ 37.48
	Men														
Noel		413 843	48.44	139.01	18.11	93.73	90.57	2983	49.40	145.77	19.27	97.59	96.37	→ 5.8 ←	↓ 3.9
Devon		516 104	56.36	143.10	17.35	99.73	86.74	3175	60.72	142.26	16.31	101.49	81.54	← 5.2 →	↓ 1.76
Taylor		576 163	62.14	201.42	27.86	131.78	139.28	3688	78.42	204.44	25.20	141.43	126.02	← 13.26 →	↓ 9.65
Jordon		239 778	61.77	160.43	19.73	111.10	98.66	1414	65.03	176.23	22.24	120.63	111.2	→ 12.53 ←	↓ 9.53
Riley		291 716	25.51	189.48	32.80	107.49	163.98	2885	76.42	135.72	11.86	106.07	59.3	← 104.67 →	↑ 1.43
	Nonbinary														
Avery		431 617	63.03	148.67	17.13	105.85	85.63	5943	78.54	167.49	17.79	123.02	88.96	→ 3.32 ←	↓ 17.17
Istus		434 190	53.45	202.55	29.82	128.00	149.1	3159	63.95	219.63	31.13	141.79	155.67	→ 6.57 ←	↓ 13.79
Alex		674 852	58.74	313.40	50.93	186.07	254.66	3806	133.26	318.58	37.06	225.92	185.33	← 69.34 →	↓ 39.85
Mello		265 426	80.79	298.38	43.52	189.59	217.59	1653	129.10	314.42	37.06	221.76	185.33	← 32.26 →	↓ 32.17
Ollie		650 404	73.27	217.77	28.90	145.52	144.51	3842	90.89	239.15	29.65	165.02	148.26	→ 3.75 ←	↓ 19.5
Kit		756 596	88.03	295.27	41.45	191.65	207.25	4545	131.71	309.62	35.58	220.67	177.91	← 29.33 →	↓ 29.02

TABLE 6.2: Comparison of Pitch values between Praat and REAPER datasets, using Median Absolute Deviation (MAD). Max(imum) and Min(imum) values are based on ±2.5 MAD to keep measurements equally expansive across groups, as REAPER's calculations exhibit a floor of 40.2Hz and a ceiling of 296.3Hz for all speakers. All values are in Hz. Note that the incredibly wide ranges recorded for some speakers (such as Remi) indicate the extremes of bi- or multimodal pitch distributions, which I discuss in more detail in Section 6.6. Full Modes/Antimodes for all speakers can be found in Appendix D.

when using REAPER data. However, in these cases where REAPER's pitch range is smaller, it is by under 15Hz or less. Larger pitch ranges are between 13.26Hz and 155.86Hz wider, indicating that REAPER tends to find overall wider ranges than Praat, given the same data.

With one exception, Praat's measurements for men are very close to REAPER's, with the difference in mean F0 measurements being less that 10Hz and the difference in pitch range under 15Hz for all men but Riley. Looking at Riley's summary statistics, we see that REAPER calculated a much larger MAD for Riley, resulting in a larger range. Among the women, Remi and Reese also show this expanded MAD in the REAPER measurements. As I examine in more detail in Section 6.5, this is an indication of interview recordings with more than one mode..

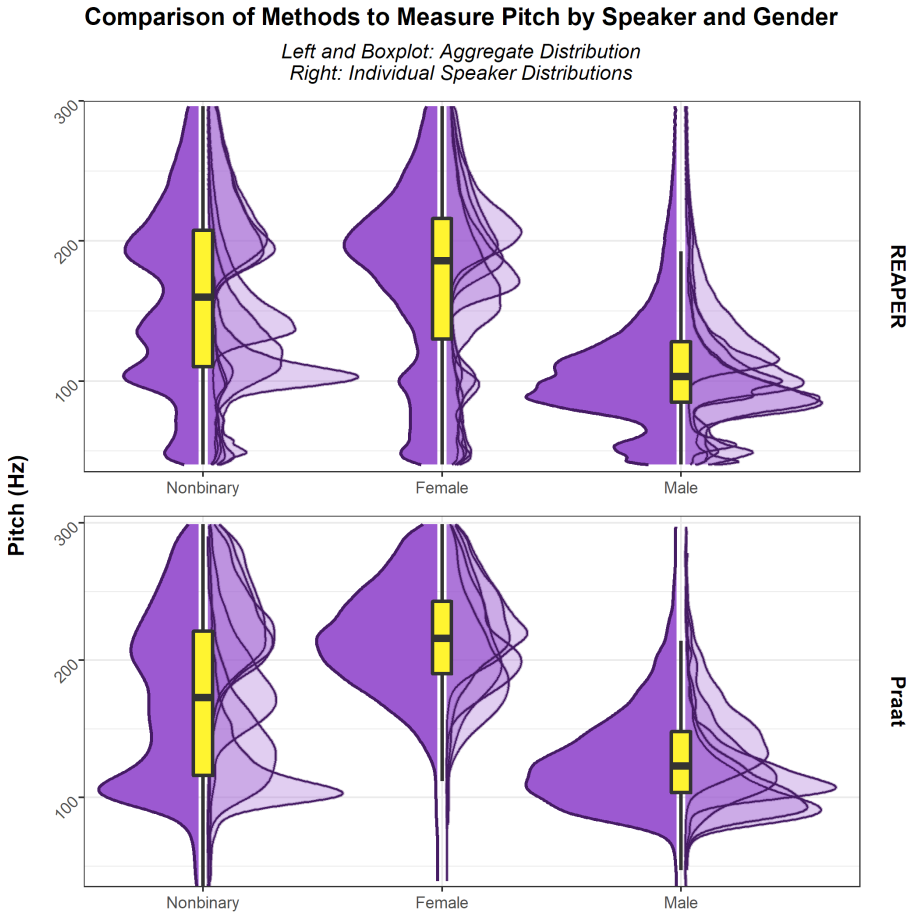


FIGURE 6.2: Comparison of Pitch Tracking with REAPER and Praat, using pitch distributions and summary statistics by gender group and individual speakers. Visualisations made using data from REAPER appear on the top, and visualisations made using data from Praat appear on the bottom. The violin plots are divided by gender group, then split in half, with aggregate speaker data on the left, individual speaker data on the right, and box-and-whiskers summary statistics of the aggregate data in the middle.

In order to visualise the similarities and differences provided by the two methods of analysis, I provide Figures 6.2 and 6.3. These two plots compare data from REAPER and Praat directly, in the form of pitch densities as measured by each method for the same recording. Figure 6.2 presents a larger, overhead view, comparing group and individual pitch densities across the two programs. While the information packed

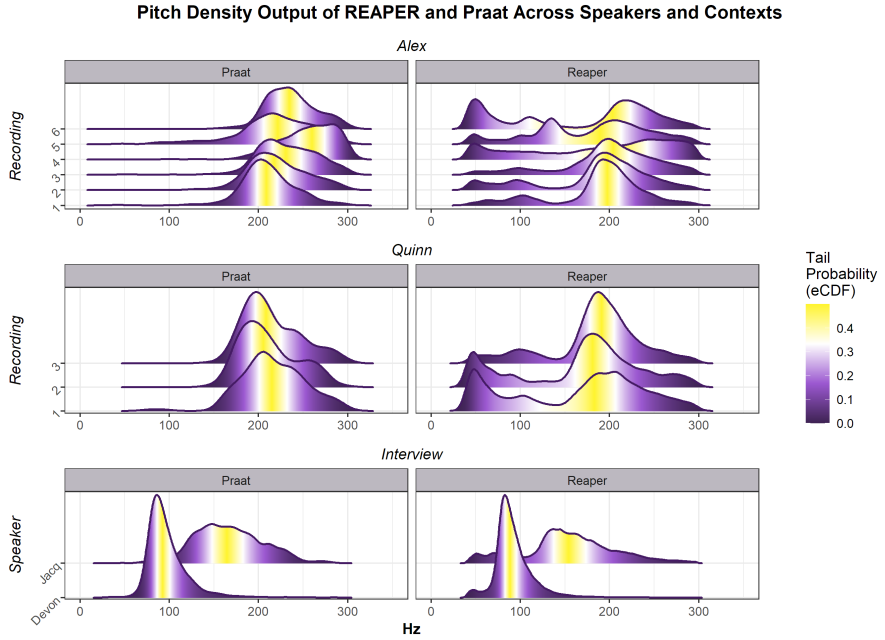


FIGURE 6.3: Density Plots comparing pitch measurements from Reaper and Praat. The yellow area band represents the central area of an empirical cumulative distribution function (eCDF), where 50% of the observed values fall on either side of the centre. It is included to visually emphasise that REAPER densities are shifted left towards the lower Hz.

into this visualisation is high, it provides an informative picture of the comparative strengths of each program, and the types of group-wide differences and similarities that are exhibited by each method. The upper plot is made using the data from REAPER and the lower plot is made with the data from Praat. Each plot is divided by gender group. The left-hand half of the violin for each gender group shows the collapsed pitch distribution, while the right half of the violin shows that aggregate data further broken down by individuals. Put another way, the low-alpha (partially transparent) data presented in the right half of the violin combines to the full alpha (opaque) distribution displayed on the left. The yellow box-and-whiskers plot visualises the summary statistics for each group. This plot provides a great deal of information: Praat tends to reduce variation across speakers of the same gender group, partially due to its trouble with creak at the lower frequencies, resulting in creak being effectively erased in the Praat data. This lack of low frequency data results in inflated pitch means, though this is more apparent for the binary speakers than the nonbinary speakers. The aggregate distribution for nonbinary people measured via Praat widens the distributions around 100Hz and 200Hz, effectively smoothing away a third, smaller peak that lies between them in the REAPER data. This has some implications in terms of interpreting potentially ambiguous incongruence from some speakers, and over-represents the similarities between nonbinary and binary male speakers as a whole, though the individual distributions for both methods do make it evident that this peak in the lower Hz is driven by a single nonbinary speaker (Avery) with a very low, very unimodal pitch distribution. Even with the high smoothing, you can still identify each individual distribution across programs, and the general/relative topographies are similar, if simplified, on Praat's side.

Figure 6.3 zooms in on a selection of recordings, one speaker from each gender

Speaker	Praat DPs	Reaper DPs
Alex	1931	448 841
Quinn	609	129 253
Devon	1861	259 313
Jacq	667	112 655

TABLE 6.3: Number of Data points (DPs) used to generate Figure 6.3. The complete number of data points for all recordings, for all participants, can be found in the "Obs" columns of Table 6.2

group, for direct comparison. Table 6.3 lists the number of data points used for each set of comparative graphs in Figure 6.3, for both Praat and REAPER. The first thing to note is, despite Praat using less than one percent of the number of data points, the general topography of both graphs is remarkably similar, especially around the main density peak for each speaker. In addition to a general lack of data in the lower range, we also see evidence of potential doubling in the Praat data, particularly in Quinn's recordings, where no measurements below 77Hz were recorded, and only 4 measurements below 150Hz were recorded[15].

6.4.4 Conclusion: Use Both

Overall, this comparative analysis confirms trends found in previous literature (Vogel et al., 2009; Styler, 2013). Praat's performance in measuring lower frequencies is generally poor, losing data even when its floor is set 10Hz lower than REAPER's (30Hz compared to REAPER's 40.2Hz), if matched with a relatively high pitch ceiling. This tendency of Praat to error out or double values leads to pitch means being inflated. An alternative interpretation here is that REAPER's measurements deflate/suppress modal pitch by incorporating too much low-frequency information from other types of phonation, such as creak. It is worth reiterating the comparison was somewhat unfair to Praat in its use of a one-size-fits-all approach to setting the initial parameters of pitch floor, ceiling, and window size, knowing this is where Praat's performance is weak. Altering these parameters, by speaker or group (essentially creating more "bins"), would definitely result in better pitch tracking, though the variation apparent across nonbinary speakers would make ay form of group-wide setting inappropriate.

The similarities between Praat and REAPER, even given their wildly different methods "under the hood", should be highlighted. Using a fraction of the data and computational power, Praat's measurements of mean pitch are relatively close to the values found by REAPER. While Praat's estimates are inflated, the inflation is consistent at around 0-10Hz for binary men and 30-40Hz for binary women. This, as well as Praat's differences in pitch range, are both due to its difficulty at accurately measuring pitch at lower frequencies. While REAPER's range measurements may be more technically accurate, what they are capturing in some cases is different phonation types (such as creaky voice). In this way, it might be better to say that

[15]Quinn's ten lowest Praat measurements (in Hz) are: 77, 94, 141, 145, 157, 157, 157, 159, 160, 162. Quinn's data was analyzed in LaBB-CAT at the same time as the other binary speakers, with a single array of settings applied to the whole group. Her total number of measurements, including measurements removed for containing errors, is in line with other speakers as well. Praat measured *minimum* possible pitch values in some of Quinn's utterances at/around floor (between 30-40Hz), so I am relatively confident that the lack of datapoints in the lower ranges here is most likely due to Praat's pitch tracking at this window size for Quinn's voice, rather than some other error causing the data to be discarded.

Praat is more suited to capturing a speaker's modal pitch range, particularly if that speaker's pitch is relatively unimodal. Ultimately, both tools are useful, not just in pitch analysis but in highlighting the shortcomings of each other, and so I have used them both in the subsequent analyses in this chapter.

Due to the truncation required in preparing files for REAPER processing, the computational load required per recording, and the preponderance of datapoints produced, REAPER's full pitch measures are unable to be used in large-scale complex analyses (without access to extensive computational resources, at least). They are also unsuitable for analyses involving specific reference to timepoints in the original, nonaltered recordings, such as tracking specific utterances, words, or phonemes across recordings. In cases where these factors are analysed, I only present data from Praat. In some cases where the scope of an analysis is limited, such as section 6.5.5 at the end of this chapter, the REAPER data was collected manually. Recall that the files suitable for REAPER analysis are converted to mono, resampled, and have extraneous information including silences outside of utterances removed. Manual collection of values in this case means physically listening to these altered sound files to find the new time selection containing the portion of interest. Where possible, pitch analyses have used the data extracted from both REAPER and Praat. I defer to REAPER, as it is the larger and richer dataset, in cases where the pitch analyses is done in isolation, without reference to time, topic, or transcription. In those latter cases, I use Praat's, and also in cases where the data from REAPER is too computationally unwieldy. Where the use of both tools can inform the analysis from different perspectives (such as Section 6.5.6), I have done so.

6.5 Analysis

6.5.1 Introduction

Having established that both means of pitch measurement are suited to different tasks, this section outlines the analysis of the data. Section 6.5.2 presents the summary statistics, pitch means, and pitch ranges. Section 6.5.3 takes a closer look at the differences in modality across speakers and contexts, while Section 6.5.4 examines pitch across recording contexts. Section 6.5.5 and 6.5.6 presents a number of Spotlights, providing a qualitative examination of a number of pitch changes within the recordings of Istus and Alex.

6.5.2 Summary Statistics, Pitch Means and Ranges

The mean pitch, minimum and maximum pitch, Median Absolute Deviation (MAD), as well as the pitch range, were obtained in R (R Core Team, 2020) using observations measured from REAPER as outlined in Section 6.4.2. These values can be seen in Table 6.4, while data for each recording (including MAD and the location and number of modes) can be found in Table D.1, Appendix D. MAD was used over standard deviation because it is more robust against non-normal distributions and outliers (Leys et al., 2013; Parish-Morris et al., 2017) and this data set contains both. As all speakers' measurements contained the effective pitch floor (40.20Hz) and ceiling (296.30Hz), the minimum and maximum pitch values were imputed from 2.5 times each recording's MAD. In cases where $\pm$2.5MAD exceeded these limits, I have listed the floor/ceiling values. I chose to use the measurable limits, rather than imputed values that cannot actually exist in the dataset, for two reasons: One, because they more clearly represent the limits of the software, and; two, because in some cases the

imputed minimums would drop below 0 Hz and provide a negative value, which is both impossible to produce and makes further measurements that depend on these values (such as range) problematic.

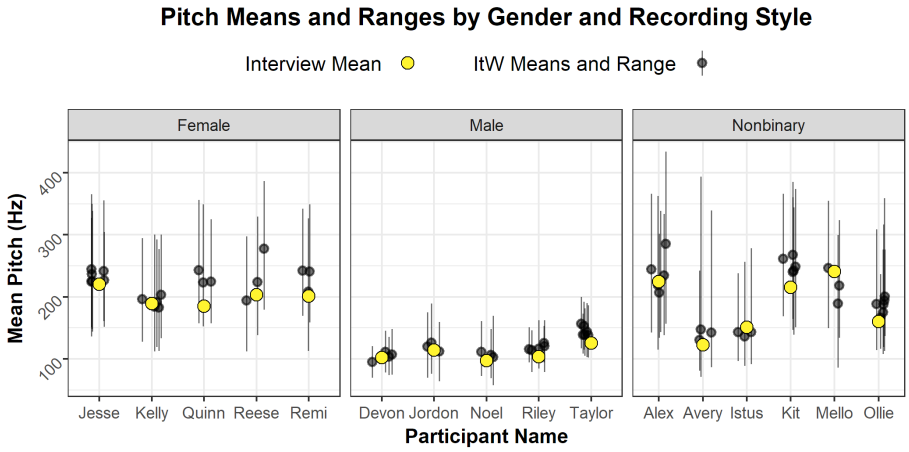


FIGURE 6.4: Pitch ranges for each ItW recording, plotted with the mean F0 of the interview. This visualisation uses the data from Praat.

Figure 6.4 plots each participant's pitch range for each ItW recording, along with their mean pitch in the interview (represented by the yellow dot), using the pitch measured in Praat. This information, as well as the summary statistics provided by REAPER (Table 6.4) confirm that women have wider pitch ranges than men across contexts, even accounting for Praat's difficulty with tracking creak. I have included some rows in Table 6.4 that are reproduced from Table 6.1 from the top of the chapter, showing the predicted measurements based on previously established binary norms. While binary mean pitch and maximum pitch are relatively similar, the minimums observed are overall much lower than found in the literature, which has a knock-on effect with the observed ranges.

While the group means reported in Table 6.4 superficially confirm Hypothesis 1, in that nonbinary participants' mean pitch and pitch range fall between the mean pitches and ranges for binary men and women, the actual story is quite a bit more complicated. Only two individual nonbinary speakers, Istus and Ollie, have a pitch mean between the group values for binary men and binary women. For all speakers, differences in cross-contextual ranges are wider within-group than across group: That is, the variation between speakers of the same gender group is equally wide across all three gender groups, and there is no obvious difference in variation between gender groups. What is unclear from both the visualisation and the summary statistics is how uniform pitch distributions are across speakers and genders, and how variations in uniformity might be affecting the summary statistics. While non-uniformity across participants is hypothesised to appear in nonbinary speakers that are producing incongruence via dissonance, it is an unexpected result in binary speakers. Fully examining this non-uniformity requires a deeper examination of the pitch distribution than the summary statistics provide.

6.5.3 Modes/Antimodes

To better elucidate the differences in pitch distribution between participants quantitatively, I examined the distribution and number of modes and antimodes in each recording. As encoded even by the terminology we use, "modal voice", pitch ranges

Gender	Name	Recordings	Observations	Mean (Hz)	Min (Hz)	Max (Hz)	Range (Hz)	C-C Range
Nonbinary								
	Alex	7	674,852	184.47	45.71	296.30	250.59	92.94
	Avery	4	431,617	106.01	40.20	173.76	133.56	13.75
	Istus	4	434,190	130.21	53.15	207.28	154.13	20.45
	Kit	7	756,596	195.88	65.55	296.30	230.75	48.62
	Mello	4	265,426	193.98	69.50	296.30	226.80	14.32
	Ollie	7	650,404	148.62	40.20	258.59	218.39	70.30
Predicted Ambiguity				160	120	200	80	
Predicted Dissonance				160	90.5	256	165.5	
Group		33	3,213,085	159.86	59.09	254.76	195.67	
Female								
	Jesse	7	478,563	194.30	105.44	283.16	177.72	27.82
	Kelly	6	595,230	164.64	44.11	285.17	241.06	26.10
	Quinn	4	232,576	169.28	49.90	288.65	238.76	40.32
	Reese	4	274,973	173.20	40.20	296.30	256.10	146.69
	Remi	4	533,401	178.93	40.20	296.30	256.10	136.77
Predicted				200	179	256	77	
Group		25	2,113,743	176.07	55.97	289.92	233.95	
Male								
	Devon	5	605,406	108.68	40.20	179.88	139.68	7.93
	Jordon	4	239,778	114.47	54.55	174.40	119.85	16.29
	Noel	4	413,843	100.38	52.93	147.83	94.90	15.42
	Riley	6	291,716	110.68	40.20	192.96	152.76	31.49
	Taylor	7	576,163	137.74	65.70	209.78	144.08	52.27
Predicted				120	90.5	165.2	74.7	
Group		26	2,126,906	114.39	50.72	180.97	130.25	

TABLE 6.4: Summary statistics for each speaker collapsed across recording, including predicted values (reproduced from Table 6.1). C-C (Cross-Context) Range refers to difference across recordings, and was obtained by measuring the primary modes of all recordings, and then subtracting the highest value from the smallest value of hese modes for each speaker. Minimum and Maximum values were imputed from ±2.5MAD, except in cases where that value exceeded REAPER's limits, in which case those limits were used (40.201Hz and 296.30Hz, respectively). The group row lists the total number of recordings, the total number of observations, and the gender group mean of the mean, minimum, and maximum values. Note that there is no group summary for the Cross-Contextual Range because the number of recordings and modes within those recordings differs for each speaker. Full data for every recording (including MAD and mode(s)) can be found in Table D.1, Appendix D.

for human speakers are generally assumed to be near-normal and therefore unimodal, with a single mode that is relatively close to the both mean and median of the distribution (as in a perfectly normal distribution, the mean and median are identical). In a distribution known to be unimodal, the mode is the most frequent measurement and can easily be visualised on a histogram. However, if a distribution has an unknown number of modes, or is known to have two or more modes, as appears to be the case given some of the pitch density distributions, determining the number and location of those modes becomes much more difficult. The multimode package in R (Ameijeiras-Alonso, Crujeiras, and Rodríguez-Casal, 2021) contains a suite of tools for calculating and visualising modality in these kinds of non-parametric distributions, using kernel density estimators (KDEs). Explaining the maths behind KDEs and their utility is beyond the scope of this thesis, but I refer to the package documentation and accompanying paper. In exploring the likelihood of multiple modes in a pitch distribution, the concept of an antimode becomes relevant. While an antimode is, simply, the least frequent measurement that falls between two modes, the presence of an antimode itself is used to determine the presence of multiple modes. KDEs identify multiple modes by identifying the presence of an antimode between them (Burman and Polonik, 2009, p. 2). In linguistics, due to the assumed near-normality of a single speaker's modal pitch, antimodes are rarely examined outside the context of creak. Within creak analyses, the lowest antimode has been used to determine the cutoff point for creaky voice (Dorreen, 2017; Dallaston and Docherty, 2019; White et al., 2022), with values below the first antimode considered to be creaky phonation and values above considered modal phonation. However, for nonbinary speakers, antimodes present at the location where we expect to find "binary" modes present an opportunity to create incongruence via dissonance. Even if all signals may be gendered according to a binary by listeners, the same cannot necessarily be said for the *lack* of a signal.

For each recording from each speaker, I first removed values under 75Hz as a rudimentary way to remove creaky voice[16]. The removal of creak from the analysis was motivated by its potential as a confounding factor: Impressionistically (as well as present in visualisations such as 6.2), participants utilised creak to different degrees. In determining whether a speaker's modal voice contains more than one mode, the presence of a "creak mode"[17] becomes problematic, especially if it is not reliably present across all speakers. While other work using REAPER has established using the first antimode as a cutoff for separating creak from modal voice (Dorreen, 2017; Dallaston and Docherty, 2019), I chose instead to apply the same cut-off to all speakers in an attempt to balance the need to discard creak with being relatively certain that *only* creak was discarded. The imperfect but universal cut-off point of 75Hz was selected to avoid cutting off any speaker's primary pitch range, given some of the participants have a mean pitch between 90-100Hz.

I used a combination of visual analysis and the multimode package in R (Ameijeiras-Alonso, Crujeiras, and Rodríguez-Casal, 2021) to analyze each recording's pitch distribution. The multimode package has functions for testing the likelihood that a set of data contains more than one mode, as well as locating and visualising any number

[16]Creak is characterised by low pitch (Keating, Garellek, and Kreiman, 2015) around 30-40Hz, but is sometimes reported to be as high as 90Hz (Laver, 1980, p. 124). While the use of creaky voice is known to be socially driven and varied by gender, and part of the motivation for using REAPER for pitch analysis was its suitability for detecting creak, a full analysis of its use by the speakers in RAINBO is beyond the scope of this thesis and something I leave to future work.

[17]Or modes. If speakers have bimodal pitch distributions, it is reasonable to assume they may have bimodal creak distributions as well

of modes and antimodes within a dataset. I first used multimode to determine the likelihood that the pitch profile for each speaker (and, later, each recording) was multimodal, against a null hypothesis that the distribution was unimodal[18]. For pitch data determined to contain a single mode (53 of the 84 total recordings), I simply located that mode. For pitch data determined to contain more than one mode, I first located 3 modes and plotted the resulting location of those modes and antimodes. When two of the three located modes were very close together (within 25Hz), I reanalyzed the mode locations and visualisation using 2 modes. Figure 6.5 contains a small sample of these visualisations, while the complete set can be found in Appendix D.

While overall most, but not all, men show strongly unimodal pitch distributions, all women have at least two modes in every recording. Nonbinary speakers do not show a strong pattern as a group, but are consistent within individuals across recordings. Two speakers, Istus and Avery, are consistently unimodal, though Avery's distributions are much narrower than Istus'. Ollie is unimodal in about half his recordings and bimodal in the other half, though his lower peak is consistently much smaller, and very close to the 75Hz cutoff. (Ollie's primary mode sits around the 150Hz mark, while his secondary peak, where it appears, is between 80 and 85Hz, just above the 75Hz cut-off for creak). Mello is mostly bimodal with a consistent peak around 210Hz but otherwise inconsistent in regards to peak and relative heights, and Kit and Alex's pitch bins are inconsistent across recordings, showing both bi- and even some *tri*-modal distributions. This presence of multiple pitch modes, and variable modes across speakers that share a gender group, including binary speakers, complicates the analysis of nonbinary productions.

6.5.4 Comparative Pitch Across Contexts

To address **H2**, I examine each speaker's pitch distribution across context. Figure 6.6 shows each participant's pitch as a smoothed density distribution, with each recording overlaid on top of each other. Pitch bins and modes/antimodes per participant, separated per recording can be found in Appendix D. While this kind of visualisation can be misleading as it gives equal weighting to each recording regardless of duration, it allows the examination of variation across contexts, and a few things become immediately apparent. Firstly, that Hypothesis 2 is refuted: Nonbinary speakers do *not* systematically present more variation across contexts than the binary speakers. While binary men do show relatively consistent pitch across contexts, the binary women, for the most part, show *more* variation than the nonbinary speakers.

There is also considerable variation between different nonbinary speakers, supporting **H3**. Istus shows more consistent pitch across recordings than any other speaker, while all of Alex's recordings are relatively distinct. Avery and Ollie show similar patterns to many of the men. Avery identifies as genderfluid and a boy, and identified as a gay dude to friends and family before beginning to ID as genderfluid. Avery's pitch distribution is the most similar to the pitch distributions of binary men, and wholly unlike other nonbinary participants, a pattern that I discuss more in Section 9.6.1. Ollie's life trajectory is a mirror in some ways to Avery's, coming

[18]According to the documentation, the `modetest` function in the multimode package has the functionality to treat any number of modes as the null hypothesis. It should therefore be technically possible to use it to determine whether a set of values contains "Two vs more than two modes", "Three vs more than three modes", and so on. However, either due to limits in computational power or to some other issue with the data or the package, no command using more than one mode as the H_0 was able to run to completion. Thus, I was unable to determine statistically how many modes (above one) some distributions contained.

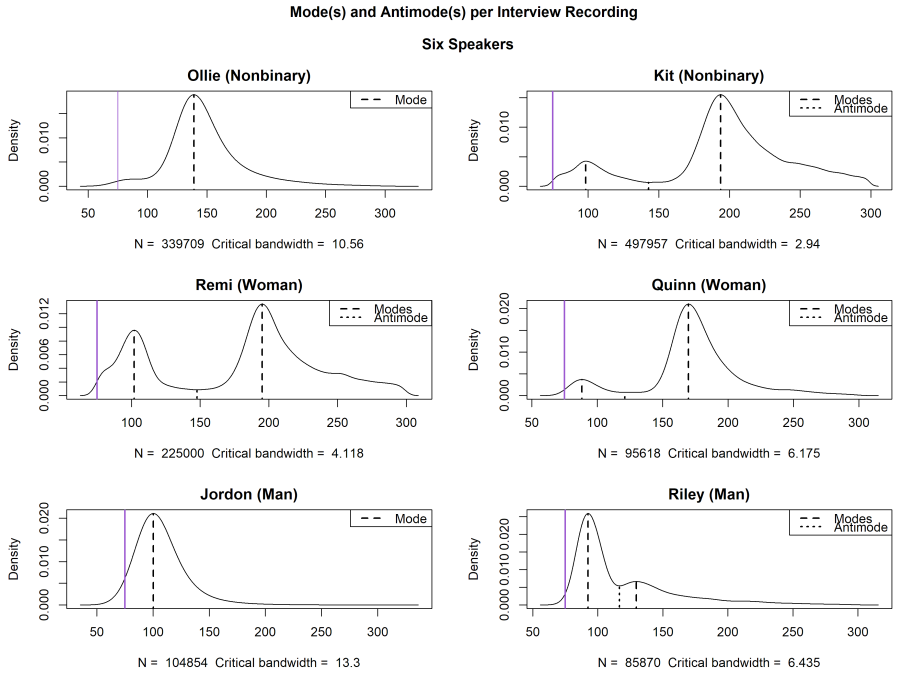


FIGURE 6.5: Modes and Antimodes of Interview Recordings from Six Speakers. The purple line on the left-hand side marks 75Hz, below which samples were removed to account for possible creaky productions. This visualisation only shows a single substantial recording - the interview - from six of the sixteen participants. Full visualisations, showing the cross- and within- speaker variation, can be found in Appendix D.

into his agender identity after considering and exploring his identity as a trans man. Alex and Kit both came into a realisation of their nonbinary identities from a starting assumption of being cishet, but twenty years apart in terms of the lived experience of what "being cishet" means. For them both, their nonbinary identity is qualified through a lens of dissatisfaction with language: Alex identifies as "nonbinary at the moment"[19], as they have not found better words, while Kit refers to their gender as being yellow in a binary world of red and blue.[20] Istus' pitch patterns are unique among all the participants, both in their position and consistency. In many ways, her productions neatly align with the predictions of incongruence via ambiguity: Her mean is higher than the average for binary men (higher than all the men in the study except Taylor), and lower than the average for binary women (and lower than each individual women, too). Her pitch range is equally central between binary productions, and consistently unimodal. I explore Istus' productions from a qualitative perspective in the next section.

Figure 6.6 also illustrates the sheer amount of complexity apparent across speakers, in addition to the within-speaker variation across context. What might be driving this within-group variation requires an analysis that is beyond the scope of this thesis in two directions: I discuss this general approach more in my concluding Chapter, 10, the strength (and weakness) of an analysis that seeks to balance the analyses of larger datasets containing multiple speakers, while still maintaining a personal connection

[19] Interview 0:0:10.

[20] "If girls were red, boys were blue, genderfluid people are purple that could be more red or more blue – I'm yellow." Interview 1:26:12.

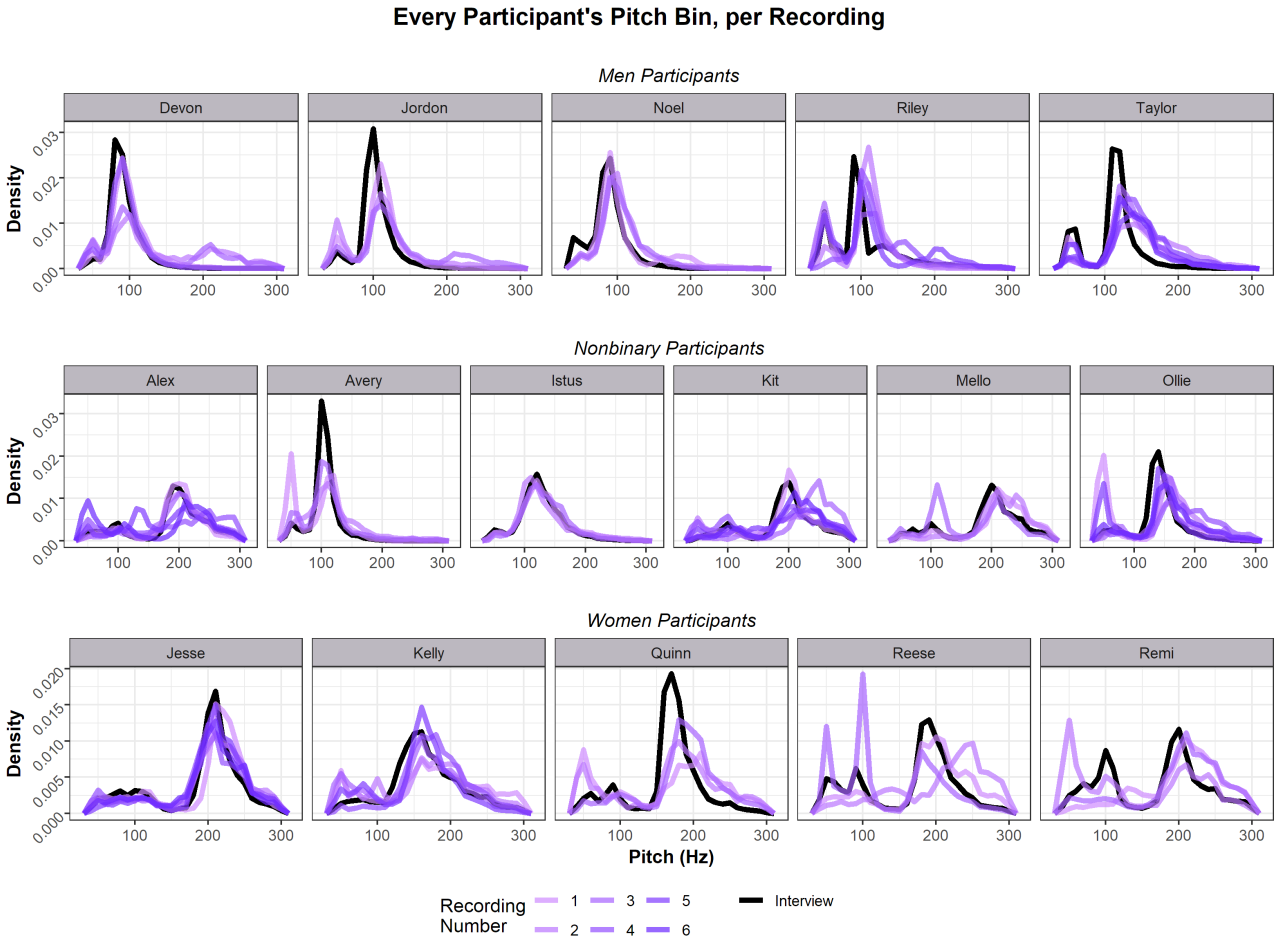


FIGURE 6.6: Pitch density distributions of every recording, by participant. Black lines represent interview recordings, purple lines represent ItW recordings. Variation across contexts can be inferred based on line difference: Alex and Reese show substantial variation across recordings, while Istus, Jesse, and Devon show very little.

to each speaker. The lack of uniformity across binary gender groups makes a more statistical approach ill-advised.

6.5.5 Spotlight: Istus

In this section, I take a qualitative approach and highlight a few excerpts from Istus' interview. Istus shows incredible understanding and control of her own speech, and specific understanding, as well, of the stereotypes involved in constructing binary masculinity and femininity across all domains. These excerpts show just how much influence speaker agency and the individual context can have in the construction of identity, and also how, in the absence of stereotypes about her own gender and the presentation of genders like hers, Istus creates something unequivocally nonbinary, and unequivocally her own.

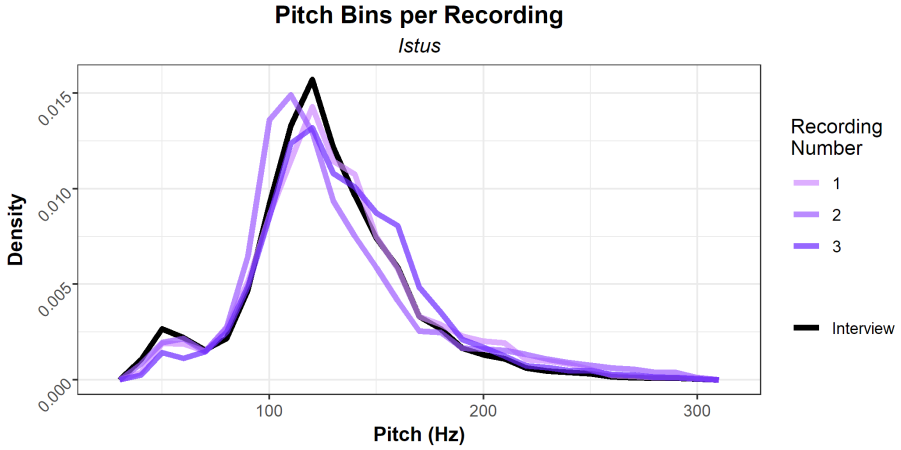


FIGURE 6.7: Istus' pitch density per recording.

Name	Recording	Observations (N)	Mean	M.A.D.	Min	Max	Range	Mode	Mode Diff. Cross-Context	Mean Diff. Cross-Context
Istus	Interview	305,360	126.66	28.85	54.55	198.77	144.23	125.00		
	1	42,110	132.09	31.38	53.65	210.53	156.89	137.93		
	2	49,868	128.86	30.60	52.36	205.35	152.99	126.98		
	3	36,852	133.24	32.49	52.03	214.46	162.43	145.45		
Summaries		434,190	130.22	30.83	53.15	207.28	154.14	133.84	20.45	6.58

TABLE 6.5: Istus' Summary Statistics, per recording and averaged. All values are in Hz, except for the Observation columns, which is N.

To better illustrate Istus' use of pitch, I provide Figure 6.7, which shows Istus' pitch distributions across recordings; it is the same data as presented within the single cell of Figure 6.6. Table 6.5 shows the individual summary statistics from each of Istus' recordings, and is the same data found in D.1, Appendix D, with some additional information added: The number of observations has been totalled, the mean of each summary statistic has been presented, and the difference between the highest and lowest mean has been presented in addition to the difference between the largest and smallest mode. Compared to the other speakers that participated in this study, regardless of gender, Istus shows the most amount of consistency in pitch across

both the interview and ItW recordings, and very little variation across contexts.[21] While she does not have the smallest cross-contextual range in terms of the difference between her highest and lowest mode, her mean for each recording is the most stable among all participants, varying by less than 10Hz. Her lowest mean is 126.66Hz and her highest mean is 133.24Hz, a difference of 6.58Hz across recordings. Both her mean pitch (130Hz) and pitch range (154Hz) fall between the group means for both binary men (mean pitch: 114Hz, pitch range: 130Hz) and binary women (mean pitch: 176Hz, pitch range: 234Hz). Istus' pitch distributions are also the most consistent across contexts, even accounting for creak (which she uses very little of). Participants with smaller differences in mode across recordings than Istus either have much smaller absolute pitch ranges across recordings (like Noel), or have more than one mode in every recording (like Mello).

Given this context, the following data is evidence of Istus' measure of control over her pitch: she knows how she wants to sound, her abilities and limitations, and through practice and observation is able to hit her speech targets consistently. I present three excerpts from Istus' interview, visualising her pitch now not as a distribution, but with the time dimension intact, to show her online pitch control, second to second. It was necessary to first reiterate her summary data across contexts to show that even the relatively small deviations in pitch presented here are actually quite outlying for Istus in terms of her own norms.

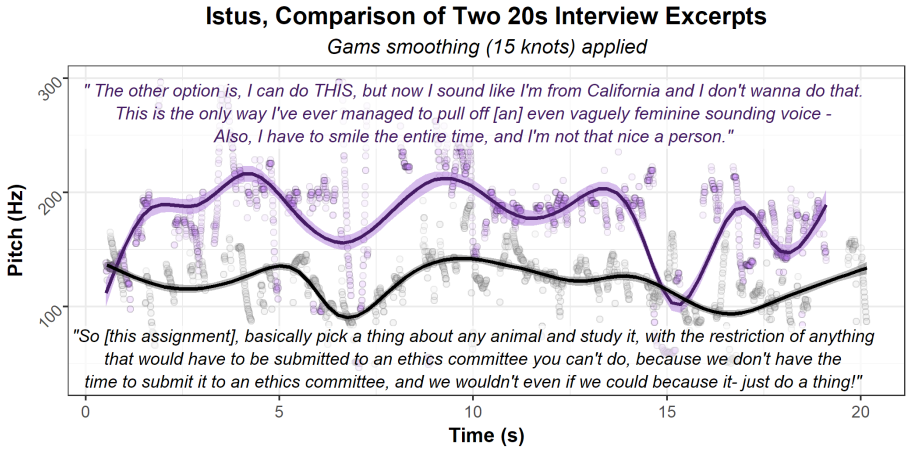


FIGURE 6.8: Two separate ~20 second excerpts from Istus' interview, overlaid on the same time dimension to show pitch differences. Both excerpts are from the same recoding, the sociolingsuitic interview. The bottom (black) excerpt is from 39:43-40:03, while the top (purple) excerpt is from 51:58-52:20. Dots represent F0 as measured by REAPER every 5 thousandths of a second where there is pitch data available.

First, Figure 6.8. This figure takes two excerpts of ~20 seconds in length from Istus' interview and aligns them on the same time dimension (with the beginning of the utterance starting just after 0s). The mean of the purple (top) utterance is 176.19Hz while the mean of the black (bottom) utterance is 120.08Hz. I have plotted

[21] To be fair, Istus' ItW recordings have relatively low contextual variation in that they all take the form of Istus telling an interlocutor about something Istus did - a school assignment, a recipe she made, etc - but the locations and interlocutors (and Istus' relationship with those interlocutors) changes about as much or more than it does with the other participants. Additionally, the style and amount of back-and-forth between interlocutors is quite different in the interview with me compared to her ItW recordings, while the pitch distribution is similar. For these reasons, I do not believe this "storytelling" genre would be the sole reason for Istus' pitch distributions to be so similar across contexts.

each individual REAPER measurement as a point, but rather than presenting a line connecting individual points, I have plotted the lines using a general additive model (gam) with a cubic spline smoother set to 15 knots[22]. In each case, I have chosen values that seem to best represent the general pitch variation I hear when listening to the audio, checked against Praat's default pitch tracking. The data in purple is Istus explicitly presenting a way of speaking she has tried using in the past to sound feminine, while the data in black is Istus describing the content of one of her ItW recordings. I selected the second (black) excerpt as being similar to the purple excerpt in duration as well as presenting a continuous narrative, while differing in that Istus is not directly speaking about gender or answering a gender-related question (thus avoiding potentially more gendered stereotypes). The purple data represents Istus presenting a "Californian" femme accent. Her pitch rises and stays high, and adopts other speech features of Californian (or Standard American) English, including the use of rhotics, a slower speech rate, and wider pitch range (Pratt and D'Onofrio, 2017, p. 288). Istus is, as she says, smiling widely during this part of the interview as well. We can see in Figure 6.8 that, before and as she is beginning to adopt the California femme voice while saying "the other option is", Istus' pitch starts in an area relatively similar to the data presented in black, and then quickly rises up to near 200 Hz and stays high (the big drop is on the end of the word "smile", the word "the", and the first syllable of the word "entire", and then pops up again on the second syllable of "entire").

In addition to her "Californian" voice, during the interview Istus describes some other vocal strategies, such as utilising a "singing style" of speech, or of talking less in situations where she wishes to be perceived as more feminine. While Istus has not visited a speech therapist, she has practiced "feminine voicing" on her own extensively, including attempting different strategies highlighting different aspects of the stereotypes surrounding what it means to "sound feminine." The results of this practice, as we can see here, is that she can drop into and out of these styles at will.

While the previous example showed an extreme pitch change, the one presented in Figure 6.9 is a much more subtle examples of Istus' control over her pitch production. While I visualised only a small portion, the more complete transcript of this section, with hesitations and speech errors edited out for clarity, is as follows:

> "[My p]itch alters sometimes, like if I'm with, you know, [a] group of men I'll, you know, speak more kind of like this. Um, *this is what my voice naturally sounds like - this is how I normally try and talk*. So like, there's a level of neutrality that I try, that I almost always try and impose but there's not much that changes too situationally I think."
>
> — Istus, Interview, 16:10.

Even listening to this section over, it is hard to specify the exact differences between the two productions (aside from the obvious uptalk at the end of the phrase, "This is how I normally try and talk."). Her pitch is indeed higher, and perceptually the second is more feminine, "softer" both in a sense of amplitude and the use of breathy voice. The first instance of "this is", in Istus' "natural" voice, sounds like there is a clear distinction between the two words, a distinction that is not maintained in the second instance during her "normal" voice. However, with the possible exception of amplitude, none of these perceptual impressions are clear from the spectrogram (see

[22]The use of smoothers and the number of knots used, both here and in the other visualisations in this section, are motivated by an intent to reduce noise given the sheer number of data points – about 100 for every second of speech.

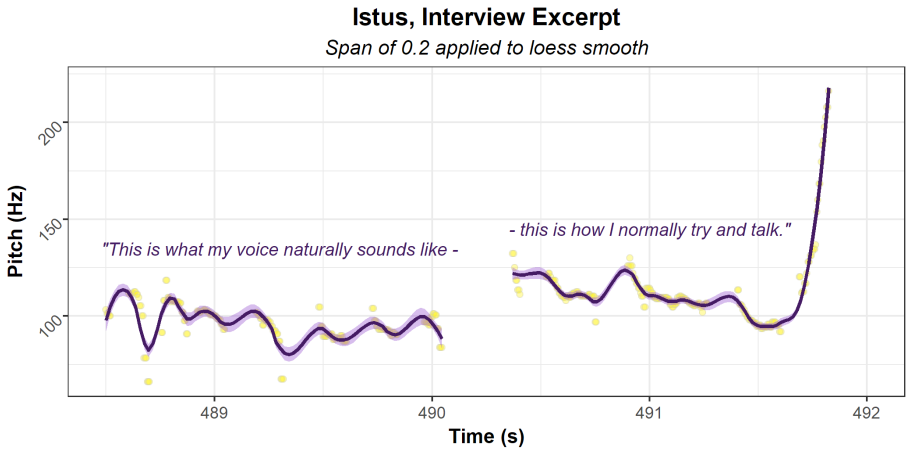


FIGURE 6.9: Excerpt from Istus' interview, at 16:10 in the REAPER file, contrasting what she calls her "natural" voice with her "normal" voice. Like Figure 6.11, this is also part of Istus' response to the question "Do you think that how you sound plays a role in how others perceive you?" Note the y-axis time labels are different to the interview timestamp in RAINBO because these pitch measurements are utilising the data from REAPER, which has all non-primary-participant speech removed and the sound re-concatenated.

Figure 6.10). The uptalk here causes some problems in comparing the mean pitch due to Praat's pitch tracking, but the mean of the "natural" utterance is 98.30 Hz, while the mean of the "normal" utterance is 114.36 Hz when uptalk is included (the mean is 108.34Hz if I cut off the word "talk" – and thus the high rising terminal – from the measurement.) This excerpt is interesting because it further illustrates Istus' control over her own production, and her awareness of multiple subtle cues to femininity, while also showing that, relative to Istus' "California" register, her pitch is actually not changing quite so drastically compared to other speech features. It is also relevant to note here Istus' terminology: Her use of "neutrality" in regards to her voice is not one that encodes a concept of over-the-top femininity (like her "California" voice), but is still explicitly moving away from masculinity, challenging the idea of the masculine as the neutral setting. The division Istus makes in the above clip is between what she refers to as her "natural" voice and her "normal" voice, and she positions her "natural" voice as neither neutral nor normal in reference to her identity.

Finally, I provide Figure 6.11. I include this figure here for a number of reasons. Firstly, because it highlights her use of humour - it is one of a number of jokes Istus makes throughout the interview that plays with gender stereotypes.[23] Her brand of humour is sometimes self-deprecating, but it is *always* self-aware, and is an integral part of the way she "does" gender in conversation. Secondly, it is an example of deliberately invoking gender incongruence via dissonance in speech, for the purpose of presenting an identity outside the gender binary – Though this particular form is exaggerated for effect and exists outside of Istus' actual speech goals. This is an identity presented not as one Istus personally holds, but rather one she adopts deliberately to make specific interlocutors uncomfortable. Finally, this excerpt illustrates that Istus' control works equally well in the direction of lowering pitch as it does in raising pitch.

[23] Istus and Kit both comment during the interview that nonbinary (and trans) people are funnier and understand not just gendered humour, but pop culture better than cis people.

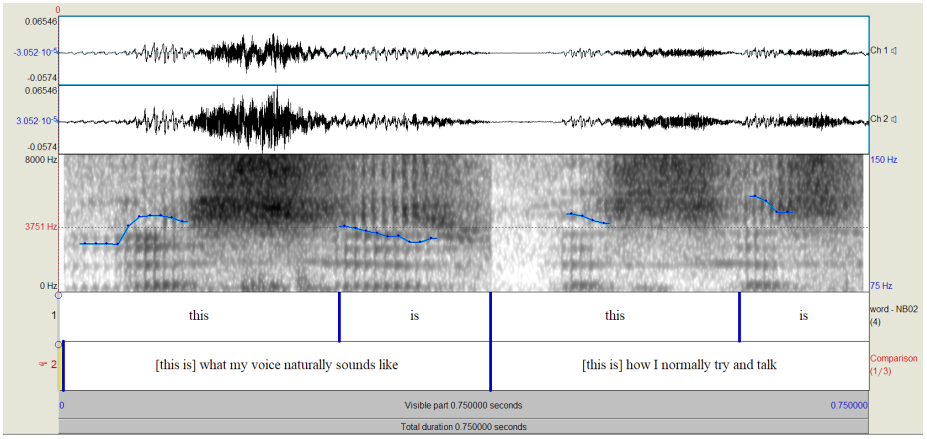


FIGURE 6.10: Textgrid Comparison of Istus' productions of "this is..." from 16:10 of her interview (the same data as presented in Figure 6.9).

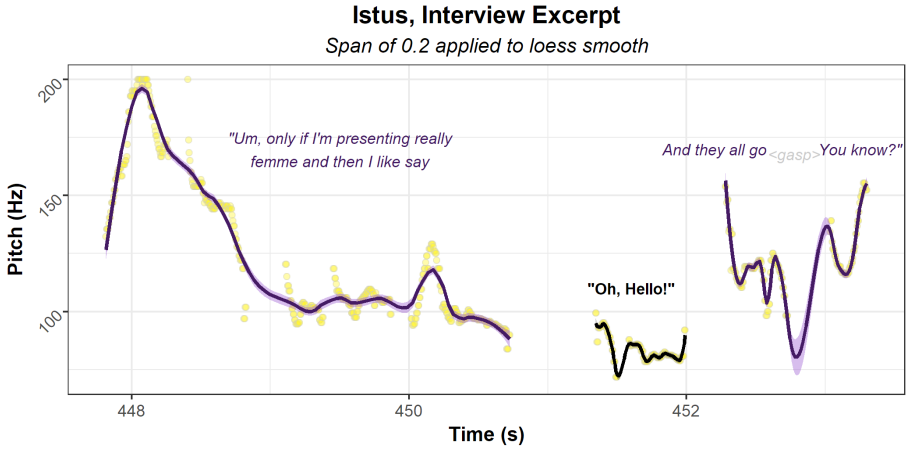


FIGURE 6.11: Excerpt from Istus' interview, at 14:53, plotting pitch over time. This contains a portion of Istus' response to the question "Do you think that how you sound plays a role in how others perceive you?". Note the trough where Istus gasps is not a drop in pitch representative of a "deep" drop in pitch while gasping, but a place where there was not enough pitch information in the data that the loess smooth extrapolated across, dragging the line towards 0.

While I have been careful to avoid broad claims about the role of agency in presenting a nonbinary identity, I think Istus represents a level of mastery of the speech domain, specialised towards gendered stereotypes in presentation. Istus' pitch measurements display incongruence via ambiguity, in that her means fall between the binary group means for both mean pitch and pitch range: Istus's mean pitch is 130Hz, higher than the mean pitch for binary men of 114Hz, and lower than the mean pitch of binary women of 176Hz; Istus' pitch range spans 154Hz, which is wider than the pitch range for binary men (130Hz) and narrower than the pitch range for binary women (234Hz). Her actual presentation and online use of pitch shows that she is hyperaware of the boundaries and is able to utilise incongruence via dissonance (as in 6.11) when she chooses, for effect.

6.5.6 Spotlight: Alex

Another nonbinary participant with explicit insight into changing the way they speak in reference to gendered norms is Alex. In this section, I present a few excerpts from Alex's ItW recordings that contain a style of speaking they identify as "Bloke Mode." Here is Alex's description of Bloke Mode, in their own words (I have edited the quote below for clarity, removing hesitations, speech errors, and repeated/filler words):

> "I've always done this thing which I've always called bloke mode, which is like... [G]oing into hardware shops and things and you go *tch~yeah! Yeah, right mate, I know what you're talking about! Yeah yeah, yeah!* And so I sometimes do that with work stuff, where I'm just trying to be ultra professional. [Bloke mode] gets into a more masculine sort of way at talking, which probably then reflects back at me, how people talk back to me."
>
> — Alex, Interview, 10:36.

I present a transcript here of Alex's Bloke Mode in action. The context for the following excerpt is that Alex has hired a younger man to complete some gardening work for them. Both speakers are moving around Alex's back garden going over the work Alex wants completed. During the interview, I asked participants for some context about every ItW recording they provide, including if they made any decisions about the way that they spoke during the conversation. Alex explicitly references the possibility that they utilised Bloke Mode in this recording[24], as they were trying to emphasise their experience, and to let the gardener know that Alex has authority and expertise regarding the work to be done and how best the young man should do it. A transcription of the recording is below, separated by utterance, with the mean pitch of each utterance listed on the right.

There are two things to bear in mind regarding these transcripts. First, due to the one-sided nature of the ItW recordings, Alex is the only speaker recorded and transcribed, so the interlocutor's input must be inferred. The line breaks follow the utterance boundaries, which themselves tend to follow the natural flow of speech, with divisions coming during pauses in the speech stream longer than ~35ms. I erred on the side of shorter utterances during transcription as the force-aligner integrated into LaBB-CAT for automatic phonemic alignment works better with smaller units. Second, the mean pitch of the utterance listed was obtained using the "Get Pitch" setting in Praat, using the same settings outlined in Section 6.4.2. Alex was one of the speakers most affected by measurement differences across Praat and REAPER, with inflated Hz measurements and missing values around the 150Hz mark (as seen clearly in the visualisations in Figures 6.12 and 6.3). Despite this, the below excerpts using Praat's pitch tracking still show the relative pitch differences apparent when Alex evokes Bloke Mode.

(1)	*Alex ItW Recording 5 0:43-1:12*	*Mean Pitch of Utterance*
1	yeah so really doing the edges	**207 Hz**
2	doing like I said that	**115 Hz**
3	round the back door	**175 Hz**
4	making sure that's a nice clear path to the back door	**142 Hz**
5	um	**113 Hz**

[24] "Um, I probably went into Bloke Mode a little bit with [person] just because, again the kind of, conveying efficiency and *this is what you need to do* and that."

6	yeah whatever	**130 Hz**
7	yeah that patch	**251 Hz**
8	um yeah	**218 Hz**
9	probably in terms of priority sort of lawn	**249 Hz**
10	back door edges that patch	**237 Hz**
11	kind of be my priority	**198 Hz**
12	list but that's	**212 Hz**
13	yeah	**240 Hz**

In (1), Alex is in Bloke Mode while they explain the areas of yard they would like the gardener to work on. While the overall pitch in this exchange is lower than some of the others presented in this section, it does not appear to be a complete and sustained shift in speech in the same way Istus presents her "normal" way of talking. It is more apparent that Alex is shifting their pitch downward on some utterances. In lines where Alex uses hedging, like line 1 ("really"), line 9 ("probably"), and line 11 ("kind of") their pitch is much higher, around 200 Hz. The utterance in this exchange with the lowest pitch is the one where Alex is reiterating their expertise in line 2 ("like I said..."), at 115 Hz.

A few minutes further into this recording, the topic of conversation has shifted away from the gardening work that needs to be done and onto international travel. I present it here as an example of an exchange without Bloke Mode invoked:

(2)

	Alex ItW Recording 5 4:02-4:56	*Mean Pitch of Utterance*
1	you probably won't get one of those in America	**237 Hz**
2	a Russian gas mask in America yeah	**325 Hz**
3	yeah	**309 Hz**
4	New Zealand's not quite as anti-Russia as America is that	**275 Hz**
5	yeah	**357 Hz**
6	yeah	**253 Hz**
7	n~never underestimate the ability of Americans to get upset by communists so	**230 Hz**
8	*laughs*	**–**
9	yeah	**236 Hz**
10	yeah	**231 Hz**
11	mmm . okay mmm	**251 Hz**
12	yeah well maybe they do I don't know I	**220 Hz**
13	it's not something I've ever really looked at when I've been in America is like	**267 Hz**
14	*laughs*	**–**
15	army surplus is not something I've really been	**268 Hz**

In this exchange, Alex is still positioning themself as having knowledge and authority, this time regarding the attitudes of Americans to Russia and communism. As this topic is removed from the immediate concerns of the work needed on Alex's garden, however, there is less need for Alex to emphasise their position as the young man's boss. Alex's authority is also less open to refute or challenge in this case, and subsequently needs less emphasis, as their expertise comes from their having been to America before. These utterances are not accompanied by the immediate fall in pitch associated with Bloke Mode. In fact, the deepest pitch drop in line 12 is expressing

uncertainty[25], and this line is still much higher than some of the utterances in (1). The average pitch of all lines in this exchange (minus lines 8 and 14 containing laughter) is 266.08 Hz. The average pitch of utterances in (1) (containing Bloke Mode) is 191.31 Hz.

For Alex, Bloke Mode is a way of speaking that uses gendered stereotypes to serve a purpose. They speak in a masculine way not to embody or present as masculine necessarily, but to convey stereotypes *associated* with masculinity that suit their conversational goals. In the recording presented here, they point to efficiency and authority as goals of the use of Bloke Mode[26]. This is an interesting contrast with Istus' California voice from the previous section: Istus originally tried to adopt the Californian voice as a way to embody and present femininity, but rejected it as inauthentic because it also embodied certain stereotypes associated with femininity (such as "smiling all the time") that Istus herself does not identify with ("I'm not that nice a person".)

6.6 Conclusions

6.6.1 Addressing Hypotheses

Hypothesis 1 and Incongruence

These results provide evidence in support of Hypothesis 1, regarding the creation of gender incongruence via either ambiguous or dissonant productions. There is evidence that Istus' pitch meets the criteria for ambiguity via incongruence, given that her pitch mean and pitch range are both firmly between the averages found for either binary gender. There is some evidence that other nonbinary speakers, such as Mello and Alex, may be creating incongruence via dissonance, as their productions consistently show two or more modes (See Figures 6.12 and D), the distributions of which fall roughly around the means for binary men (114 Hz) and binary women (176 Hz). However, these findings are complicated by the data from the binary women, who also show multimodal tendencies in this same direction, and the overall variability of the presence/absence of multiple modes for all gender groups.

Hypothesis 2 and Contextual Variation

Hypothesis 2 is refuted. As shown in Figure 6.6, binary women exhibit much higher pitch variation across context than either male or nonbinary speakers. Some nonbinary speakers, such as Alex and Kit, do show variation across contexts on par with the binary women, and one binary woman (Jesse) is quite stable across contexts. However, these are exceptions. Further, in all cases, the variation found between individuals within the binary women or nonbinary gender groups is wider than the variation across gender groups. While the data opposes the predictions made by **H2** regarding the influence on context on speakers, it also provides a number of opportunities for reflection, and directions for future inquiry. The stability across contexts exhibited by some speakers, like Istus and Jesse, point towards a degree of intention in production that was not accounted for by **H2**. The underlying assumption of **H2** positions nonbinary speakers as somewhat unsettled in their identity

[25] Also, this line is delivered more softly than the surrounding speech, with a lot of creak, and comes across as directed more towards the recording equipment or Alex themselves than the person they are speaking to

[26] A few years on, discussing Bloke Mode, Alex pointed to this recording as also being about the ladnscaping conversation about conveying "traditionally masculine subject knowledge".

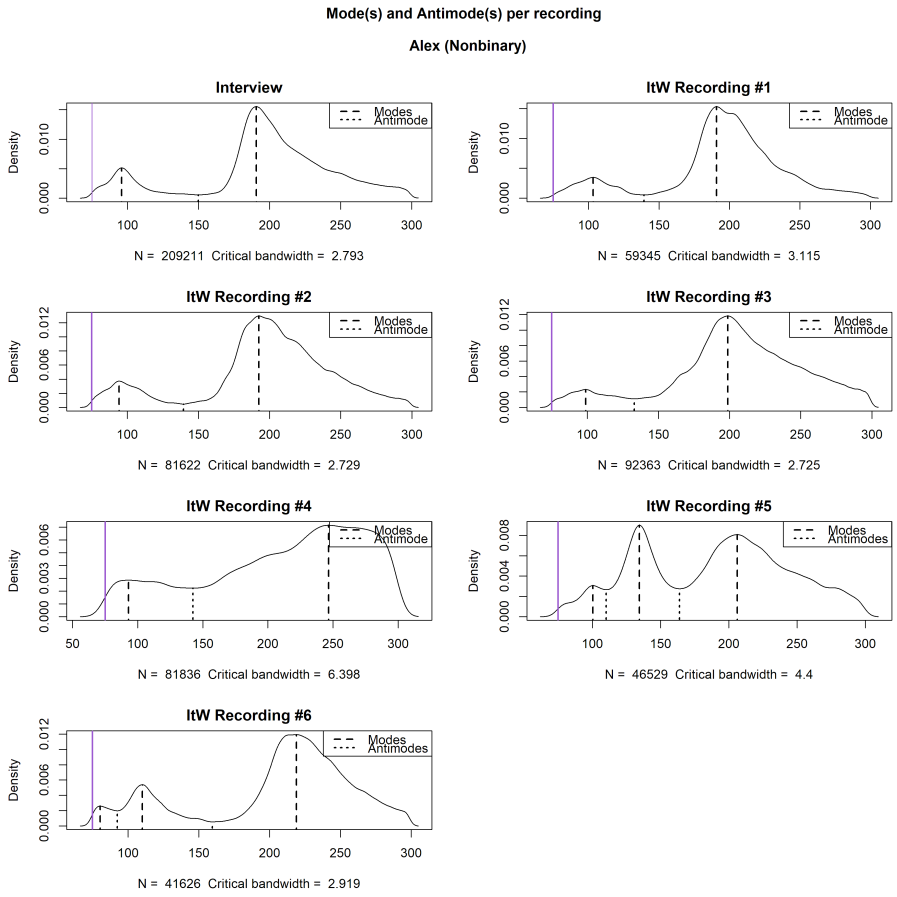


FIGURE 6.12: Modes and Antimodes from all of Alex's recordings. The cell labelled ItW #5 is the one the transcript is taken from, where Alex is speaking to a gardener about some work they would like completed in their back yard. This recording also clearly shows a change in pitch distributions, in line with Alex's Bloke Mode. (Note that ItW #4 looks odd because Alex is chatting with two young children that are very excited.)

relative to binary speakers, and assumes this manifests as being particularly prone to reactive performance, rather than enactive production. I return to this line of inquiry in Chapter 10.

Hypothesis 3 and the Individual Condition

Hypothesis 3 is clearly confirmed when examining pitch. This is illustrated qualitatively when comparing Istus and Alex, who produced speech unlike that of either the binary men or the binary women, but utilise very different strategies (and discuss very different goals) for doing so. This is also clear in the raw data, such as in Figure 6.6, where we see stark differences in pitch across nonbinary participants. The influence of the individual condition can also be seen among binary participants, particularly the women, though there is still more similarity across the binary same-gender-group members in pitch means and distributions than the nonbinary ones, and not just in a way where individual nonbinary members can neatly be slotted as patterning with binary men or binary women exclusively.

6.6.2 Conclusion

This chapter explored the hypothesis that nonbinary speakers may treat a gradient variable, like pitch, in novel ways as part of presenting their gender. There was some evidence of pitch being used in presenting gender incongruence, though it was not present in all nonbinary speakers, and it appeared that the stability of this presentation was stronger for some participants, like Istus, compared to others. Further, some evidence for the incongruence hypothesis (**H1**) was presented in terms of examining the overall pitch distributions of some participants exhibiting incongruence via ambiguity. A confounding factor was identified in presenting evidence that pitch ranges across women (and nonbinary speakers) may be estimated as wider when exploring summary statistics, due to narrower distributions around multiple speech modes rather than a continuously wide range across a single mode. **H2** is refuted in that the women show wider variation across contexts than men and nonbinary speakers. While technically nonbinary speakers' contextual variation was between that found for binary women and binary men, variation between individual nonbinary speakers (and individual women) is much wider than the variation across these two gender groups. **H3** is upheld in examining the differences between participants that share gender groups, and in participant's online changes in pitch in conjunction with their own insights about their speech production and speech goals. The was upheld with qualitative examinations of speech in conjunction with their own insights and discussion of speech goals, such as Alex's Bloke Mode and Istus' "California voice."

Of the three broad categories of acoustic variables under consideration (along with vowels and frication), pitch represents the most obvious starting point in an examination of gendered speech. However, this analysis has shown that even the most obvious gendered stereotypes in speech contain a wealth of complexity and points requiring some amount of methodological reflection. The comparison of Praat to REAPER showed that even if the broad trends are the same, there is nuance in the lower F0 that is often lost when using an established tool like Praat. The broad strokes in which pitch range is often gendered was also challenged, complicated by the presence of multiple modes within the productions of single speakers, indicating a rich vein of future inquiry into the examination of gender and contextual effects on pitch range**s**. Having started with an examination of the obvious, and finding a wealth of unexpected complexity, I turn now to something less overtly stereotyped, but nonetheless strongly implicated as gendered within linguistics: vowels. As pitch is assumed to be gendered as a consequence of speaker physiology, vowel production is assumed to be gendered as a consequence of the social factors influencing sound change.

Chapter 7

Vowels

7.1 Introduction

This chapter examines differences in vowel production based on speaker gender. It is divided into two main sections, addressing monophthongs (Section 7.2) and Vowel Space Area (VSA) (Section 7.3). In organising this chapter, I begin my analysis with more concrete measures, and move towards relatively abstracted ones, and begin with monophthong production in isolation across speaker groups and context. I then examine the integrated formants (particularly F2) of the monophthongs of interest, and then the entire vowel system in tandem via an exploration of the VSA. I present VSA second as it represents a somewhat more abstract category and, because of this, has a tendency to be rooted more closely in the literature to generalisations regarding the physiological consequences of gender. Finally, in Section 7.4, I take a small deviation from the main hypotheses regarding incongruence to spotlight the older triad's productions of NEAR and SQUARE with reference to previous literature on New Zealand English (NZE) involving speakers of similar years of birth.

7.1.1 Defining and Measuring the Vowel Space

As discussed in Chapter 2, formants are the acoustic consequence of the interaction between the source (the vibration of the vocal folds) and the filter (the vocal tract). Pitch refers to the way the fundamental frequency (F0) is perceived, while vowel perception in speech is primarily driven by F1 and F2. Formant production depends on both relatively static features of the body (such as the thickness of the vocal folds or length of the vocal tract), as well as more malleable features such as the position of the tongue back in the mouth.

In speech production, a single vowel is generally interpreted as existing at some point in an abstracted acoustic space made up of the F1 measurement on one axis and the F2 on the other. These formant frequencies can be abstracted further to correlate roughly with the position of the tongue relative to the rest of the mouth: The tongue body's relative height corresponds to F1, where a low F1 correlates to a high tongue body, and a high F1 correlates to a low tongue body. The relative saggital position of the highest point of the tongue correlates roughly to F2, where a high F2 is regarded as a fronter vowel with the highest part of the tongue closer to the front of the mouth, and a low F2 is regarded as a backer vowel with the highest part of the tongue closer to the back of the mouth (Ladefoged and Johnson, 2014, p. 22). This malleability is core to speech as a communication system, as it is this manipulation that allows speech sounds to happen. The relationship between formant frequencies and tongue height is one reason the vowel space is often plotted obliquely, with the 0 point in the top right of a chart.

If one plots every vowel from a given speaker on a graph like this, and draws a convex hull around some or all of the outermost vowels, then the VSA can be calculated as the area inside this convex hull. In the past, VSA has often been calculated from simpler shapes, taking only three, four, or five "cardinal" vowels thought to make up the widest points of the vowel space for all speakers of a given dialect (Jacewicz, Fox, and Salmons (2007), though see Fox and Jacewicz (2008) for a comparison of smaller quadrangles to a "total vowel space area" measurement from polygons individually calculated per speaker). I go into more detail in section 7.3.4 exactly how VSA was calculated in this study, but suffice it to say that modern computational power makes measuring the VSA from multiple vowels relatively trivial, and I took the approach of getting the most accurate hull possible.

7.2 Monophthongs

Section Outine

This section first presents the state of monopthongs in NZE, according to the current literature. In Section 7.2.1, the predictions made by the hypotheses are presented. Section 7.2.2 and 7.2.3 outline the analysis and results for addressing **H1** and **H2**, while Section 7.2.4 presents the individual Spotlight data, used in addressing **H3**. After a discussion of the results in Section 7.2.5, I conclude in Section 7.2.6.

Background: NZE Vowels

NZE is uniquely suited to the examination of an emerging dialect due to its relatively recent development, and the existence of historical recorded speech samples dating from first-generation European immigrants through to contemporary times under the umbrella of the Origins of New Zealand English (ONZE) Project (Maclagan and Gordon, 2004). Previous and current analyses using these corpora – the Mobile Unit, The Intermediate Archive, and the Canterbury corpus – present the development of contemporary NZE as emerging from multiple British regional dialects with some influence from Australia (Gordon et al., 2004, p. 37) in the latter half of the 19th century (Gordon et al., 2004, p. 102). Over time, NZE has experienced a push-chain shift in the short front vowels DRESS, TRAP, and KIT (Langstrof, 2006, p. 142; Langstrof, 2011, p. 138), with TRAP and DRESS being raised and fronted, and KIT centralising in response. This shift continues to the present, to varying degrees, among speakers (Przewozny and Viollain, 2016), with FLEECE the most recently implicated: FLEECE is starting to diphthongise in response to the overlap/encroachment of DRESS (Maclagan and Hay, 2007).

Sound change is assumed to be led by women when that change is not overtly socially proscribed (Labov, 1994), and that adage has tended to hold true in examinations of sound change in NZE. The speed and degree of the push-chain shift across speakers and time has been found to be predicted by word frequency (Hay et al., 2015), word age (Walker and Hay, 2011), speaker age, socioeconomic status (Maclagan and Gordon, 1996; Maclagan and Hay, 2007), individual "leaders" and "laggers" (Brand et al., forthcoming), and, indeed, gender. Studies that examine the historical

gender effect consistently find women are further along the shift than men.[1] That is, higher and fronter realisations for TRAP and DRESS, and more central realisations for KIT (Gordon et al., 2004, p. 211; Woods, 1997). There is an assumption that gender differences in the production of singular vowels, and hence the factors involved in sound change, are primarily driven by the social choices made by speakers in manipulating the vocal tract, and that changes to all vowels in a system are assumed to be driven by the relatively static forces mentioned above. However, differences in production are not the entire story: Drager (2011) also found that listener's *expectations* regarding speaker gender and age play a role in the perception of DRESS and TRAP. This is in line with literature I previously covered in Section 2.2.1 on the negative impact that violations of listener expectation have on gender classification and intelligibility (Ryalls and Lieberman, 1982; Bradlow, Torretta, and Pisoni, 1996; Diehl et al., 1996). These findings have implications for any work involving nonbinary or otherwise "atypical" (expectation-violating) speakers. As nonbinary participants must navigate mismatches between their identity and others' perceptions of them as a matter of course, work examining their production alone, as this chapter does, is incomplete.

7.2.1 Monophthongs: Hypotheses

Vowels represent vast possibilities for categorisation, and the following predictions about incongruence vary depending on the way vowels are parsed and gendered by nonbinary speakers. This section examines eleven NZE monophthongs[2] in isolation, first normalised across speakers using a modified Lobanov approach, and then using unnormalised raw values. Normalising is traditional when studying vowels in isolation, particularly in the context of (binary) gendered differences in sound change, where the goal is to abstract away from the system-level effects presumed to be caused by physiology, which itself is presumed to be asocial. Conversely, by analysing unnormalised values across binary and nonbinary populations, the extent to which these presumed-"innate" physiological changes are manipulable can be interrogated. Within the unnormalised values, I examine the monophthongs as singular discrete units made up of a combination of F1 and F2 measurements, as isolated F1 and F2 measurements, and in formants across the whole vowel system. I treat each monophthong (or formant) as a single gradient variable, and therefore the predictions borne from **H1b** apply. While I examine multiple signals across variables in Chapter 9, I spend some time in this chapter examining multiple vowels in tandem, as there are different possible outcomes predicted depending on whether the whole vowel space is simply the product of incongruence within single vowels, or treated as a domain for multiple signal incongruence (and therefore under the consideration of Hypothesis **1C**).

Hypothesis H1b – Gradient Signal Incongruence *Given a single gradient variable, nonbinary speakers can communicate identity by producing incongruence either via* **ambiguity**, *in the form of more centrally-distributed tokens compared to binary male and binary*

[1] Socioeconomic Status (SES) also plays a major role, with lower-SES men being more advanced in the changes reported by Maclagan and Gordon (1996), though authors like Woods (1997, pp. 99–100) point out that there are some inherent categorical issues in the classification of SES in Aotearoa New Zealand. I did collect SES information from participants, in the form of their own and their parents' occupation and level of education. However, with very few exceptions, mostly between members of the older triad, and across members of the older triad the younger participants, the participants in my study are essentially homogeneous in terms of SES. I have tried to make note of cases where specific differences in SES between individual participants may be relevant, but for the most part I do not consider it further within this thesis.

[2] DRESS, FLEECE, FOOT, GOOSE, KIT, LOT, NURSE, START, STRUT, THOUGHT, and TRAP.

female speakers, or via **dissonance**, *by producing tokens in a bimodal fashion, where a single speaker's productions are within both binary male and binary female ranges.*

H1b Within Single Vowels

I provide a visualisation of incongruence arising via ambiguity using simulated data in Figure 7.1. In this case, when a given vowel's productions are examined, a nonbinary speaker utilising ambiguous gender incongruence will produce a range of tokens in between the production clouds produced by men and those produced by women. The simulated data presented in this section are intentionally extreme for the sake of illustration, and it is expected (and indeed was found) that gendered differences in measured vowel production exhibit much greater overlap. Because of this overlap, testing **H1b** involves first establishing that the acoustic properties of a given monophthong (or formant) are distinguishable between men and women.

Further, it is not expected that all vowels produced by a single nonbinary speaker will exhibit incongruence equally. Gendered variation in vowels is most likely to be found more strongly in vowels representing extreme tongue articulations (FLEECE, GOOSE, and TRAP) (Pépiot, 2009; Weirich and Simpson, 2013; Coleman, 1976). As the normalisation process I use (outlined in more detail below) scales F1 and F2 together, differences of degree across F1 and F2 are predicted only in unnormalised data, and it is in the unnormalised data that F2 is predicted to vary more strongly than F1 across all vowels. In the absence of external influencing factors such as the effects of a sound change in progress, the literature predicts the widest variation between binary participants to be in the F2 of FLEECE. Following this, I predict that nonbinary speakers will manifest incongruence more strongly in F2 over F1, and in peripheral/extremely articulated vowels over more central ones.

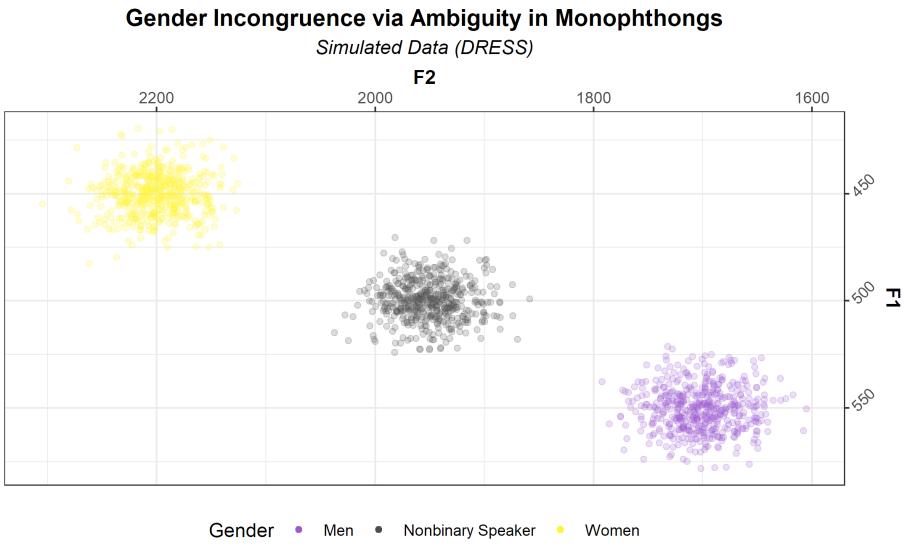


FIGURE 7.1: Comparison of a single nonbinary speaker's distribution of tokens of a single vowel (in this case DRESS, represented by the grey points) compared to the distributions of men (purple points) and women (yellow points). Incongruence via ambiguity is achieved as the nonbinary speaker's production cloud is somewhere between the expected values for men and women.

Incongruence arising via dissonance, conversely, is visualised in Figure 7.2. In this case, a given nonbinary speaker's productions are within the expected values for *both*

men and women, with very few tokens appearing in the central space. Incongruence in this cases arises from apparently-mutually-exclusive productions appearing from a single source.

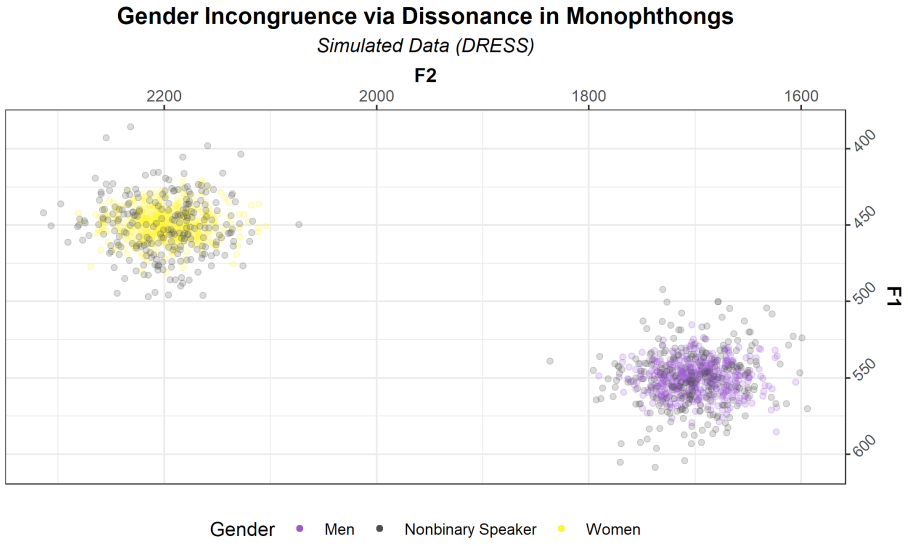


FIGURE 7.2: Comparison of a single nonbinary speaker's distribution of tokens of a single vowel (in this case DRESS, represented by grey points) compared to the distribution of men (purple points) and women (yellow points) when incongruence is achieved via dissonance. The nonbinary speaker is producing vowels in bimodal distribution, whose peaks approximate the expected values for both men and women.

H1b Across the Whole Vowel System

In trying to determine the exact ways a vowel is interpreted by listeners along gendered lines, a number of possibilities arise. Some monophthongs may be gendered as single acoustic targets that fall along a gradient, some may be gendered differently between formants, and still others may only be gendered in consideration to the entire vowel space. However, trying to account for all these possibilities per speaker causes the prediction-space to balloon outrageously. For this reason, I limit the visualised predictions artificially by assuming that all vowels are treated with a similar scope by a given speaker.

Figure 7.3 represents the expected composition if each monophthong in the system is producing gender incongruence equally via ambiguity or dissonance, respectively. These would merely be the consequence of single ambiguous/dissonant productions and do not themselves represent some consideration of the entire vowel system on the part of the speaker. While I have no reason to expect speakers would utilise different types of incongruence for different vowels, the resulting vowel space if some are produced utilising dissonance and some are produced utilising ambiguity is easy to imagine as a combination of the plots presented in Figure 7.3.

Hypothesis H1c – Multiple Signal Dissonance – *Given multiple signals, nonbinary speakers can communicate identity by producing incongruence across signals.*

Evidence from chain shifts show that it is possible, at least in theory, that nonbinary speakers may treat the entire vowel space as a cohesive unit in terms of presenting gender: In a chain shift, a vowel infringing on the space of another can cause the

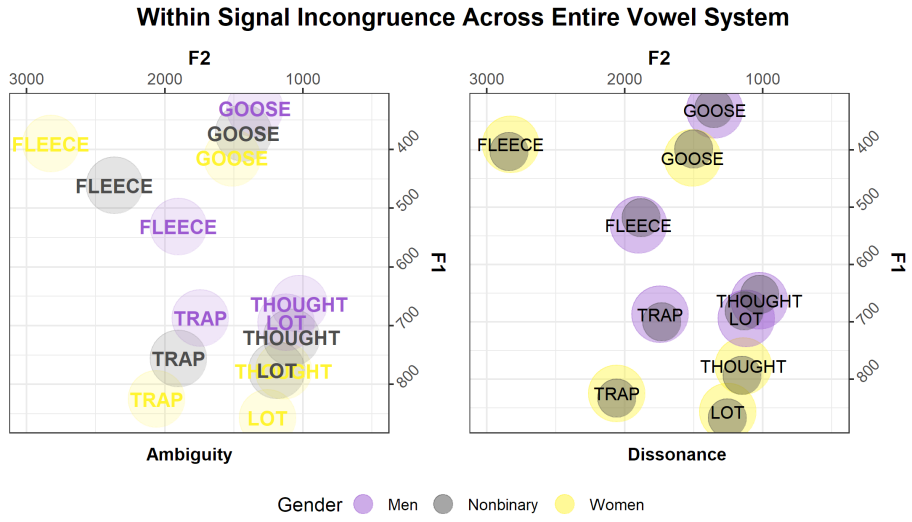


FIGURE 7.3: Expected presentation of a single nonbinary speaker's vowel system, if incongruence was created equally within each vowel via ambiguity (on the left) or dissonance (on the right). The ambiguous visualisation represents an expansion of Figure 7.1 across the whole vowel system, while the dissonant visualisation represents an expansion of Figure 7.2.

second vowel to move away from the infringing one, which can, in turn, cause a shift in a third vowel, and so on. If a speaker is aware of "masculine" and "feminine" production targets for each vowel, multiple-signal dissonance can be achieved via producing some vowels within feminine expectations, and some within masculine expectations. One possible manifestation of this is visualised in Figure 7.4. Given the differences in frequency, salience, and perceptual weighting, it is unexpected that all vowels within a system will be treated equally, but this does not necessarily indicate multiple-signal incongruence. For example, studies of sound change show us that young women are more likely to use innovative forms, and in NZE specifically, Maclagan and Hay (2004) found that some young women produced DRESS higher than FLEECE. Given this, it would be expected that innovative realisations of DRESS (higher) and FLEECE (lower) carried a heavier association with femininity than other vowels (such as STRUT) that are not directly implicated in this specific variation. All of this is to say that variation between the relative size of the peaks in a dissonant (bimodal) system can likely be attributed to differences in these weightings, though this would be a difficult thing to parse for every vowel for every speaker, even if the data were abundant.

Hypotheses 2 and 3

Hypothesis 2 – *Immediate contextual factors have a greater influence on nonbinary speakers than binary speakers.*

H2 predicts that nonbinary speakers will produce greater variation across contexts than binary speakers. This is tested by looking at vowel production across different recordings per speaker and measuring the amount of variation found.

Hypothesis 3 – *Variation is bound by the individual circumstances of each participant. Individuals with shared socially-encoded circumstances, such as age and binary gender, will speak more like each other.*

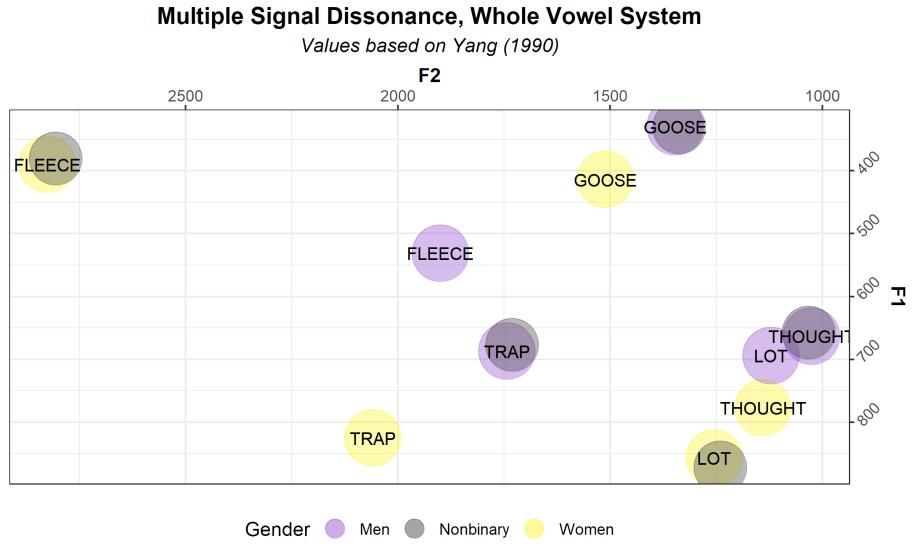


FIGURE 7.4: Expected presentation of gender incongruence across a whole-vowel system. Should a nonbinary speaker be utilising this type of incongruence, singular vowels examined in isolation are expected to align more closely with the prototypical productions of men *or* women, while an examination of all vowels will show variability between "masc-aligned" and "femme-aligned" vowels. Incongruence here is created via different vowels patterning with different genders.

H3 predicts less variation overall in binary gender groups as compared to the nonbinary gender group, as not all nonbinary speakers will create incongruence in all variables, or create incongruence in similar ways, and that each speaker's productions will be tempered by each speaker's individual condition.

Should the binary men and women in this sample be found to differ significantly from each other, **H3** predicts those differences will follow the directions previously established, due to greater similarities in the individual condition between same-binary-gender participants. While full attention to the myriad factors identified to influence production (such as word frequency, word and speaker age, leader/lagger status, and other sociodemographic variables) is far beyond the scope of this chapter, it is expected that, should differences be found among the binary speakers in RAINBO, these should be more similar within binary-gender groups than across them, and in the same direction as found previously in NZE: Women should be more advanced along the high-front vowel shift, and gender variation should be more observable in vowels with relatively extreme articulation (such as FLEECE and GOOSE), and along F2 in comparison to F1 (with consideration given to the NZE context of TRAP/DRESS raising and KIT centralising).

7.2.2 Monophthongs: Analysis

Timestamps for all transcribed monophthongs were extracted using Fromont and Hay (2012)'s Language, Brain & Behaviour Corpus Analysis Tool (LaBB-CAT) at the segment level with a window offset of 0 seconds. A Praat script automated through LaBB-CAT's web interface then extracted the F1 and F2 at the midpoint of each vowel. Tokens matching the following criteria were removed, following the process outlined in Brand et al. (2019): missing F1 or F2 measurements; F1 or F2 measurements over or under 2.5 times the standard deviation for each speaker; a duration shorter than

0.01 seconds or longer than 3 seconds; those appearing inside unstressed syllables or before an [l]; interrupted speech or hesitations; appearing in words within utterances whose speech rate was over 11 syllables per second; and stopwords[3]. After pruning, two further removals were made based on the paucity of remaining data. These were tokens of stressed schwa across all participants, and the remaining measurements of participant Mello. Overall, after outlier removal, the number of tokens was reduced from 114,588 to 31,577. Final token counts for each vowel and gender group are listed in Table 7.1.

Vowel	Gender			Total
	Nonbinary	Women	Men	
DRESS	3313	1218	1197	5728
KIT	2904	1046	1070	5020
STRUT	2239	800	825	3864
FLEECE	2181	835	773	3789
LOT	1981	697	826	3504
TRAP	1584	498	595	2677
THOUGHT	1190	474	410	2074
GOOSE	801	412	413	1626
START	689	262	306	1257
FOOT	589	237	241	1067
Total	18 030	6659	6888	31 577

TABLE 7.1: Number of tokens of each monophthong remaining for analysis after pruning.

Analyses were run on both normalised data and raw data. Normalisation was achieved using a modified implementation of the Lobanov formula developed by Brand et al. (2019, p. 13). In standard Lobanov normalisation, F1 and F2 values are normalised based on the means and standard deviations of all vowels per speaker, which can be problematic if there is a high variability in token counts across speakers and categories (as is the case in the present data set: The speaker with the least number of vowel tokens has 612, while the speaker with the most has 4999). In Brand et al. (2019)'s implementation, means are taken from *each* vowel for each speaker, and the data are normalised based on a mean of those means (the standard deviation across means is also calculated, rather than a general standard deviation across all vowels).

Mixed effects linear regressions were run for the F1 and F2 of each vowel to determine if differences in formant production are meaningfully linked to speaker gender. An overall effect of nonbinary gender is not predicted by **H1**, unless all speakers happen to be producing incongruence via ambiguity strongly enough to differentiate their productions from binary men and women. However, this modelling is useful in identifying the general position of nonbinary speakers with respect to men and women, as well as the productions of men and women with respect to each other and past research. Mixed effects models are so-called because they contain both fixed

[3] a, ah, ahh, am, an, an, and, are, aren't, as, at, aw, because, but, could, do, don't, eh, for, from, gonna, had, has, have, he, he's, her, high, him, huh, I, I'll, I'm, I've, I'd, in, into, is, it, it's, its, just, mean, my, nah, not, of, oh, on, or, our, says, she, she's, should, so, than, that, that's, the, them, there, there's, they, this, to, uh, um, up, was, wasn't, we, were, what, when, which, who, with, would, yeah, you, you've. Note this wordlist cleaning was not perfect and a handful of stopwords with alternative spellings did appear in the final data set.

and random effects. Random effects are non-repeatable sources of variance assumed to arise randomly from a given population, such as WORD, since the exact words that appear in the recordings and the number of times each word appears are different in each recording. Fixed effects are those which were tested and controlled for, and considered to be a potential cause of the variance observed, such as participant gender (coded here as TRIGEN, to indicate it is a three-way division between men, women, and nonbinary participants). The random effects were the participant (coded as SPEAKER) and the word being said (coded as TEXT). The random effect of SPEAKER controls for consistent variation among individual speakers, and the random effect of TEXT controls for consistent variation of individual words. These are random effects because I can predict that they will influence production, but do not consider them meaningful when examining the influence of gender on overall vowel production in the context of **H1B** (though there is some overlap between SPEAKER as a random effect and the individual condition considered in **H3**). The results of the mixed effect modelling regression guided future analysis and visualisation, in indicating which vowels presented the greatest amount of variation across gender groups.

7.2.3 Monophthongs: Results

Hypothesis 1b: Individual Monophthongs and System-wide F2

Running the models on normalised vowels resulted in no significant effect of gender on vowel production. If the effect exists at all, it may be too small to be detectable in the limited number of participants and data points in this study.[4]

Results of the linear regression on the unnormalised data can be found in Table 7.2 (normalised results data can be found in Appendix E). Given the statistical power of these models is already low due to the dearth of tokens, I treat p as a source of guidance rather than a binary measure, as suggested by Halsey et al., 2015. While some, including Jekel (1977) and Halsey et al. (2015), recommend not reporting p-values at all, instead I report *all* p-values between 0 and 0.099. I err conservative regarding p-values between 0.04 and 0.06, and discard them all as relatively weak evidence of difference. For p-values below 0.04, I take the suggestion of Dahiru (2008), Kirk (2001), (and others), and report both the estimates and confidence intervals for these models, as found in Table 7.3. In all cases, the estimate of nonbinary speakers was found to be between that of binary men and binary women. With the exception of the F2 of FLEECE, nonbinary speakers were found to produce significantly lower Hz than binary women, but did not vary significantly from the productions of men.

[4]There is another possibility: The lack of significant variation may be a result of the convergence of men's and women's vowels over time. To attempt to address this, I turned to the Canterbury Corpus archive, one of the three archives that make up the Origins of New Zealand English project (Gordon, Maclagan, and Hay, 2007) to examine the likelihood of finding no significant differences across the entire vowel space, given a similar sample. I used a bootstrapping method where I examined just over 1000 subsets of speakers from the Canterbury Corpus that were similar to my sample of speakers (5 men and 5 women, born within 5 years of each other). Of these 1067 iterations, only 3% (37) resulted in a model with no significant gender differences in vowel F1 and F2. However, within iterations, as the sampled speakers became younger the number of significantly different vowels fell abruptly. This supports a convergence hypothesis, given the Canterbury Corpus only contains data for speakers born up until 1988, while all but three of my participants were born after 1997. However, even beyond this gap in birth-year data, there were other methodological and technical issues with my bootstrapping approach: Uneven years-of-birth across the population resulted in oversampling for some date ranges, and issues with sampling without replacement in others. As the examination of diachronic vowel changes among binary speakers is only tangentially within the scope of this thesis, rather than continue further down this fascinating rabbit hole, I abandoned this thread of inquiry and do not address it beyond this very long footnote.

Model Results				
	F1			
	p	Intercept	Est. (Gender)	Sig.
START	0.0732	739.74	87.56 (W)	No
THOUGHT	$p \geq 0.1$	473.51		No
TRAP	$p \geq 0.1$	615.58		No
DRESS	$p \geq 0.1$	435.62		No
FLEECE	0.0845	404.532	48.059 (W)	No
KIT	$p \geq 0.1$	497.73		No
LOT	$p \geq 0.1$	605.81		No
GOOSE	$p \geq 0.1$	408.16		No
FOOT	0.0787	485.03	47.35 (W)	No
STRUT	0.0920	679.52	73.63 (W)	No
	F2			
	p	Intercept	Est. (Gender)	Sig.
START	0.0516	1280.34	128.88 (W)	Weak
THOUGHT	0.0181	881.97	155.23 (W)	Yes
TRAP	0.0448	1798.45	199.63 (W)	Weak
DRESS	0.0569	1999.58	174.93 (W)	Weak
FLEECE	0.0559	2129.88	-148.10 (M)	Weak
KIT	0.0200	1620.76	118.08 (W)	Yes
LOT	0.0209	1060.60	124.896 (W)	Yes
GOOSE	0.0551	1716.35	135.69 (W)	Weak
FOOT	$p \geq 0.1$	1321.26		No
STRUT	0.0146	1311.58	153.21 (W)	Yes

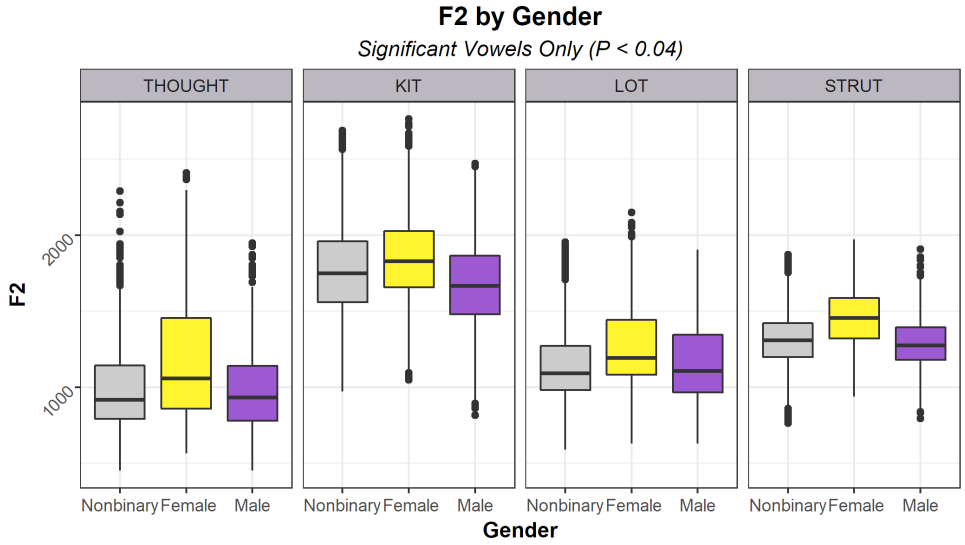
TABLE 7.2: Results of multiple linear mixed models, examining the relationship between formants and gender in vowel production, using raw (non-normalised) data. Results from normalised data can be found in Appendix E. Models in R took the form *VowelFormant~trigen + (1 | speaker) + (1 | text)*.

Regarding the F2 of FLEECE, nonbinary speakers were found to produce significantly higher F2 than binary men, but in this case they did not vary significantly from the women. In no case do nonbinary speakers produce a vowel significantly differently along F1 or F2 than men *and* women.

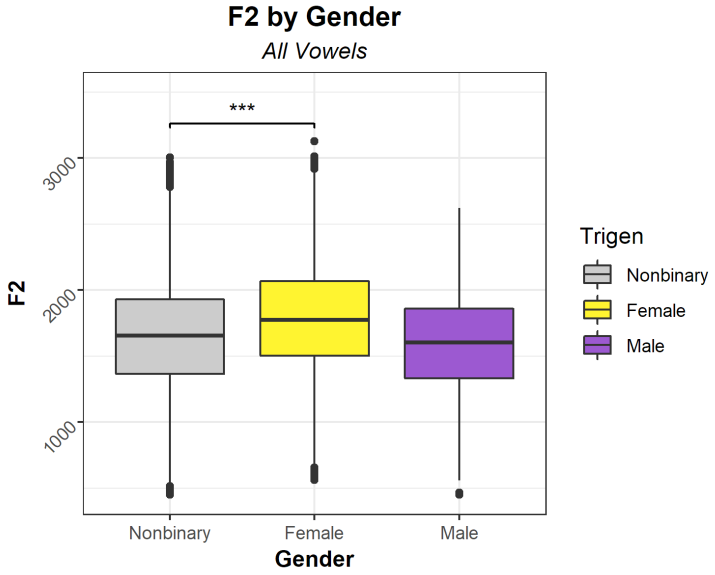
Four models show strong evidence ($p < 0.04$) of differences in F2 as a result of speaker gender; THOUGHT, KIT, LOT and STRUT. Five models show weak evidence ($0.04 \leq p \leq 0.06$); TRAP, START, DRESS, FLEECE, and GOOSE. In all cases the trend is the same: Nonbinary productions of F2 as a group fall somewhere between the productions of men and women. With the exception of FLEECE, where nonbinary speakers' F2 is higher than men's but indistinguishable from women's, all cases with significant $p \leq 0.06$ are in the direction of nonbinary speaker's productions being lower than women's, while not being higher than men's.[5]

To further explore these intriguing F2 results, I constructed another linear model to examine the effect of speaker gender on F2 only, with VOWEL as an additional fixed effect with 10 levels, and random intercepts for speaker and word. The results of this model found that F2 varied significantly as a result of speaker gender ($\chi^2(2)$=15.185,

[5]In the raw data, there is overlap across individuals and gender groups for all vowels.



(i) Box plot of F2 by gender for THOUGHT, KIT, LOT), and (STRUT).



(ii) Box plot of F2 of all vowels by speaker gender.

FIGURE 7.5: Boxplots of F2 production by gender group among vowels with significant variation (7.5i), and all vowels (7.5ii).

Vowel	Intercept	Estimate	CI
THOUGHT	881.97	155.23	49.156 - 261.265
KIT	1620.76	118.08	34.853 - 201.292
LOT	1060.60	124.90	36.460 - 213.382
STRUT	1311.58	153.21	52.310 - 254.105

TABLE 7.3: Results of modeling difference in F1 and F2 across vowels, using unnormalised data. Nonbinary speakers serve as the reference term in these models, and estimate provided are the difference between nonbinary speakers and binary women.

p=0.0005043), with women's productions 132.5 Hz $\pm$ 42.58 (standard errors) higher relative to nonbinary speakers and 190.70 Hz $\pm$ 42.63 (standard errors) relative to men. Nonbinary speaker productions were significantly lower than womens productions, but were not found to vary significantly from men, though it is the case that the F2 productions of nonbinary speakers for all vowels fell between that of men and women. The results of this modelling are visualised as box plots in Figure 7.5, both with the modelled data using all vowels included in the model (7.5ii), as well as with only those four vowels found to show significant variation in previous modelling (7.5i).

I present the F2 information divided by speaker and vowel in Figure 7.6. As the number of data points for each vowel for each speaker varies widely, I divided them into 20 quantiles, so each dot represents 5% of the available data for that vowel and speaker. While there is very little uniformity within any gender group, it is particularly notable that the nonbinary speakers do not simply fall into a bimodal distribution, whereby some speaker productions are identical to the binary men and others are identical to the binary women. I talk about this more in Chapter 9 when examining multiple signals.

To summarise, **H1b** addressed three possibilities, in normalised vs unnormalised data, and in examining individual vowels vs whole-system effects. In examining individual monophthongs, no gender effect was found in normalised data, while unnormalised data found no effect of gender on F1, but in F2 identified a relatively strong gender effect in THOUGHT, KIT, LOT, and STRUT, a marginal gender effect in START, TRAP, DRESS, FLEECE and GOOSE, and no gender effect in FOOT. In examining F2 production in all monophthongs, normalised data showed no significant differences. In examining the unnormalised data, the F2 of nonbinary speakers is between that produced by men and women, while modelling indicated an effect of gender on production whereby women produce significantly higher F2 than men and nonbinary speakers, while men and nonbinary speakers are not distinguished statistically.

Hypothesis 2: Vowels in Context

To address **H2** and explore the influence of context, separate linear mixed effects regression models were also run on non-normalised data for each speaker to determine whether there was an effect of recording (as a stand-in for interlocutor/context) on the production of F1 and F2 of each vowel, treating word as a random effect. I do not report the full output of the 144 models here, but Table 7.4 lists the proportion of significant differences for each formant across contexts as a percentage per formant per speaker, and for speaker and gender groups overall. For each speaker, the number

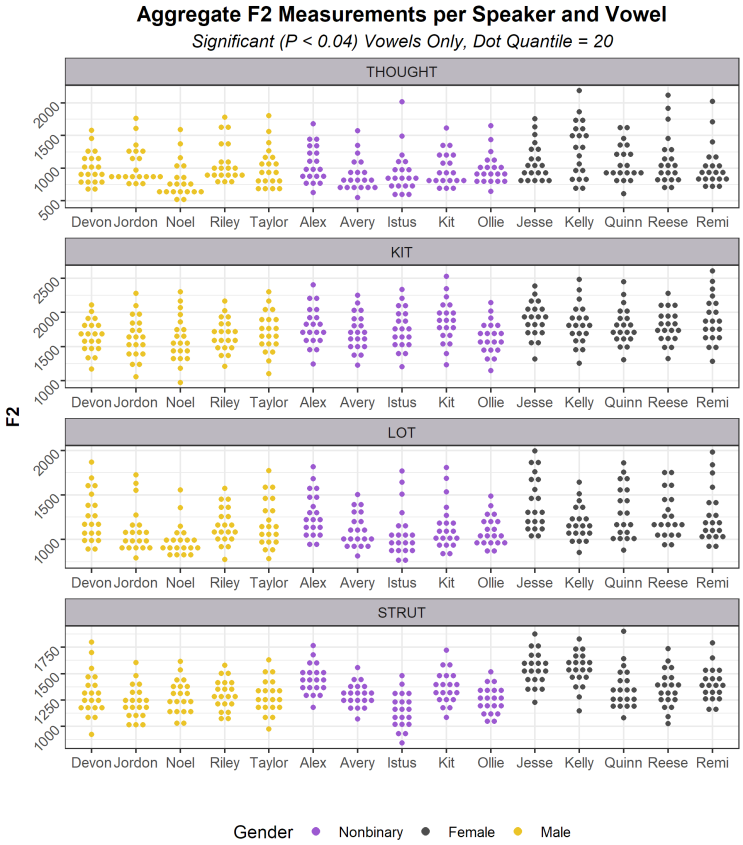


FIGURE 7.6: Dot quantile plot of F2 by speaker and vowel. Due to differences in the amount of tokens in each group, data was converted to quantiles: Each dot represents 5% of the total measurements for each vowel and speaker.

of predictors was between 40-140[6]. Aggregating these results shows that, as a group, nonbinary speakers vary in production significantly across contexts about 24% of the time, compared to women who vary 18% of the time, and men who vary 23% of the time. Within each gender group, there are individuals of all genders that exhibit a high amount of contextual variation across both formants, individuals with a low amount of variation, and individuals whose variability differs by formant. In the latter case, variability is always higher in F1 than F2, with only two speakers, Istus and Noel, exhibiting more significantly varying vowels in F2 than F1, by 5% and 3.33%, respectively.

7.2.4 Monophthongs: Spotlights

This section highlights some interesting comparisons and insights underlying the individual data. I present only a handful of visualisations in this section, but have included full visualisations for every vowel for every speaker in Appendix F.

[6]Number of transcribed recordings - 1 (intercept) x 10 (vowels) x 2 (formants), - (1, 2, or 3) per speaker as some low-frequency vowels did not appear in shorter recordings at all.

Speaker	Significantly Varying Vowels (as a percentage of total)			Percentage by Gender
	F1	F2	Combined F1 & F2	
Alex	33	13	23	
Avery	43	7	25	
Istus	0	3	2	
Kit	34	30	32	
Ollie	39	19	29	
			Nonbinary People	24.09
Devon	24	14	19	
Jordon	65	20	43	
Noel	5	10	8	
Riley	33	15	24	
Taylor	30	18	23	
			Binary Men	22.81*
Jesse	14	12	13	
Kelly	25	20	23	
Quinn	5	0	3	
Reese	30	10	20	
Remi	50	20	35	
			Binary Women	18

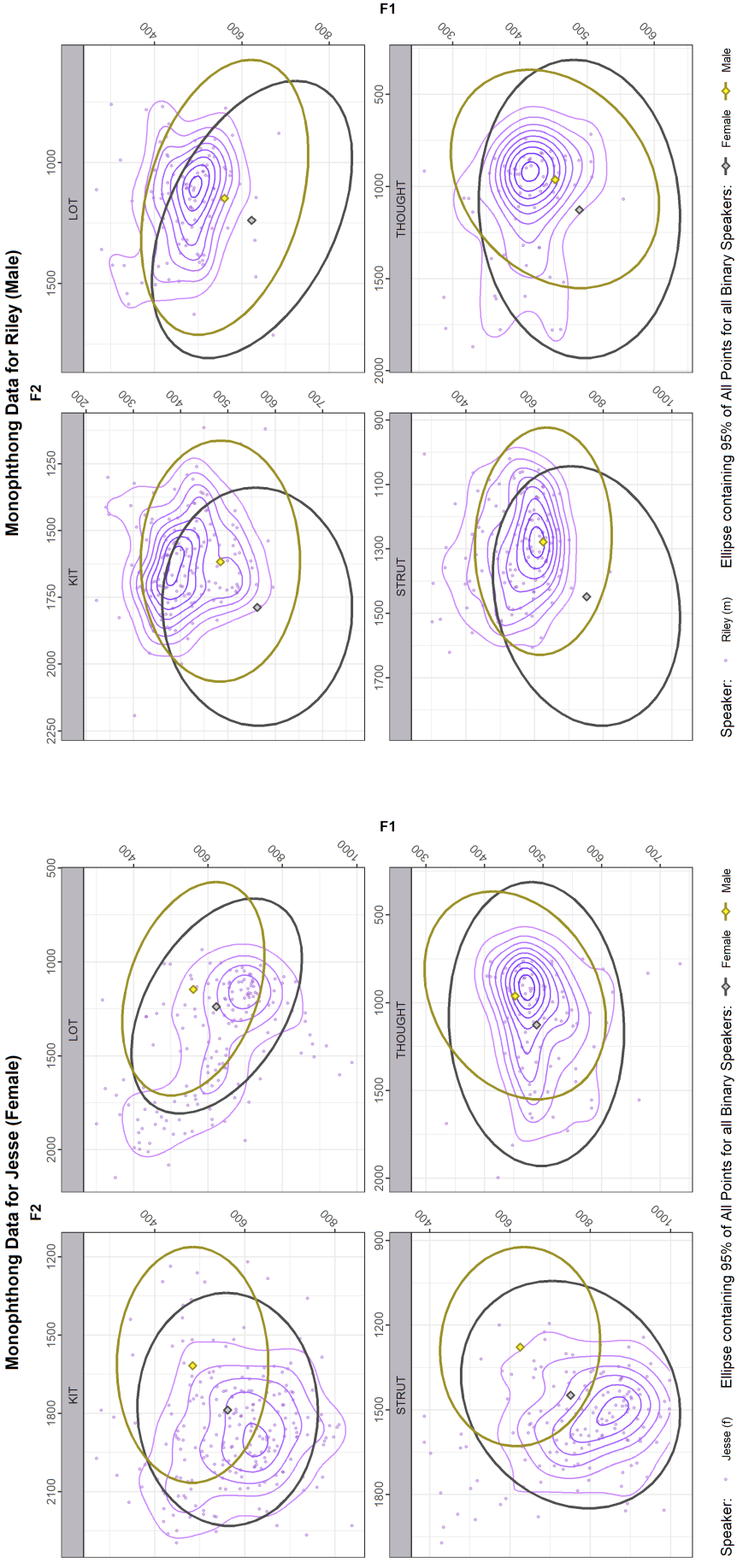
TABLE 7.4: Proportion of Statistically Significant Differences in Monophthong F1 and F2 Across Recordings. Summaries of the context of each In the Wild (ItW) recording can be found in Appendix C.
** This drops to 20% if the single recording where the speaker is intoxicated is removed.*

Spotlight: Jesse & Riley

To illustrate the binary norms that nonbinary speakers are working within, I begin with a brief examination of two of the binary speakers. Figure 7.7 shows the vowel productions of Riley (a binary male) and Jesse (a binary female), compared to all binary speakers in the data set. Monophthongs shown are those whose models resulted in strong evidence of difference ($p < 0.04$) in F2 across gender. I chose these two speakers as examples of prototypical production within the expected binary vowel space. Most binary speakers follow a similar pattern, with some exceptions: Noel's data points tend to cluster around F1 and F2 values that are much lower than the other male norms, and Reese's STRUT vowel is very similar to the distribution for men. All visualisations for all speakers can be found in Appendix F. Overall, binary speaker productions fall largely within the ellipses representing the binary gender they identify as. Of course, this is tautological in that the speakers themselves make up the group ellipses, but this is why I chose to visualise Jesse and Riley as representative. Kelly (a binary women) and Taylor (a binary man) are much closer to the data ellipses of their gender, but they are also the two speakers with the largest number of data points: Kelly's tokens make up about 30% of all tokens for binary women (2029/6659), while Taylor's are about 43% (2982/6888).

Spotlight: Alex & Kit

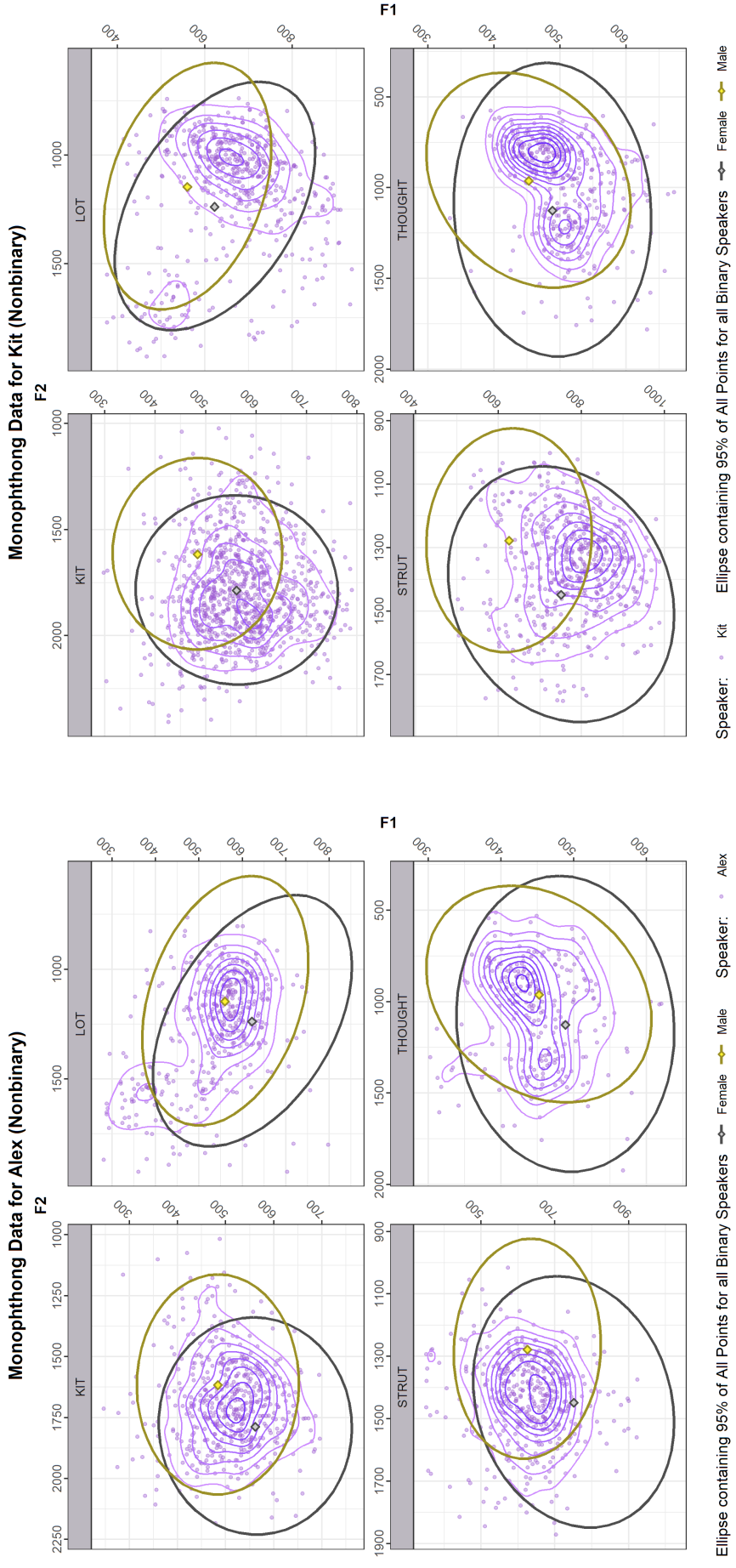
Having established the binary production space and how binary speaker's data constructs (and falls within) it, I turn to the nonbinary speakers. Figure 7.8 illustrates



(i) N = 226/1046 (KIT), 159/697 (LOT), 147/800 (STRUT), 84/474 (THOUGHT)

(ii) N = 148/1070 (KIT),100/826 (LOT), 127/825 (STRUT), 65/410 (THOUGHT)

FIGURE 7.7: Distribution of selected vowels from Jesse, a binary woman (7.7i), and Riley, a binary man (7.7ii), plotted alongside data ellipses representing 95% of all tokens for binary men and women (Jesse & Riley included). Note that the diamond represents the simple center of the data ellipse and not the mean or mode of the actual data. Monophthongs shown are those whose models resulted in significant differences ($p < 0.04$) in F2 across gender. Contour lines are set to display 10 "bins" - approximately 10% of tokens are within each contour. Note that Riley and Jesse's data falls largely within the ellipses representing the binary gender they identify as. Token counts for each vowel, as well as the total number of tokens for that binary gender, are listed under each plot.



(i) N = 500 (KIT), 412 (LOT), 431 (STRUT), 249 (THOUGHT)

(ii) N = 967 (KIT), 517 (LOT), 544 (STRUT), 384 (THOUGHT)

FIGURE 7.8: Distribution of monophthings from nonbinary participants Alex (7.8i) and Kit (7.8ii), plotted alongside data ellipses representing 95% of all tokens for binary men and women. Monophthongs shown are those demonstrating significant differences ($p < 0.04$) across gender. Contour lines are set to display 10 "bins" - approximately 10% of tokens are within each contour. Alex's vowel distribution is aligned with the hypothesised predictions regarding within-signal incongruence via ambiguity, while Kit's vowel distribution is aligned with the hypothesised predictions regarding within-signal incongruence via dissonance. Token counts for each vowel are listed under each plot. Note the bimodal distribution of THOUGHT for both speakers is likely an artefact of the centralisation/reduction of *sort* in the context of the hedge <*sort of*>.

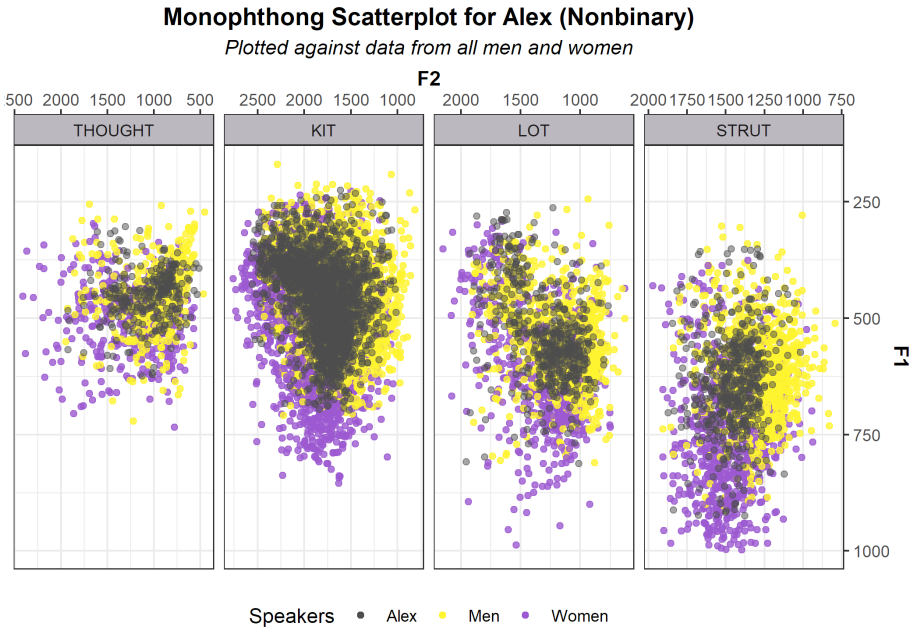
two nonbinary speakers (Alex and Kit) who appear to be creating incongruence via different strategies in vowel production, consistent with the predictions outlined in **H1b**. Alex's productions (7.8i) are centrally located between men and women, especially along the F2 dimension where modelling identified a significant likelihood of difference across gender. Alex's productions are also relatively dense, and nearly all their tokens fall within the area of overlap between the data ellipses for men and women. Conversely, Kit's tokens are much more widely distributed across the overall vowel space along both F1 and F2, spanning the areas of both men and women. The dual rings of the THOUGHT vowel for both speakers is mostly driven by the centralisation/reduction of *sort* in the context of the hedge <*sort of*>. Both speakers show a similar intention across vowels – that is, whether the outcome is incongruence via ambiguity (for Alex) or via dissonance (for Kit), the strategy remains the same for every vowel in the system, resembling the two "perfect" theorised approaches to **H1b** proposed in Figure 7.3, albeit with some expected real-world noisiness.

Spotlight: Alex & Avery

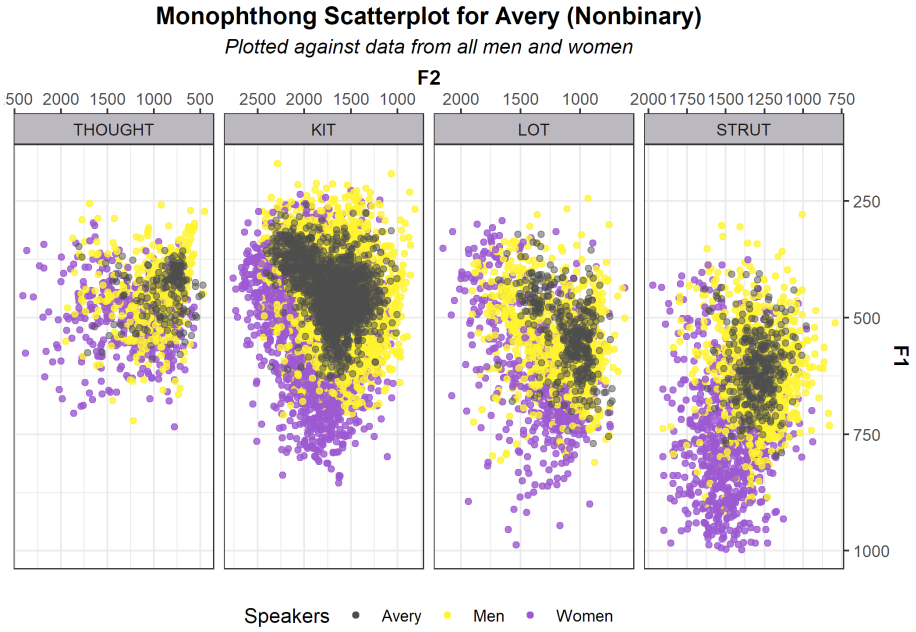
Figure 7.9 plots the data from two nonbinary speakers, Alex and Avery, against the aggregated data from all binary men and women. I have deliberately modeled it to match the original idealised visualisations earlier in this chapter (Figures 7.1 and 7.2). Comparing Alex (top) and Avery (bottom), we can see that Alex's productions are more centralised between men and women, while Avery's vowels are relatively concentrated within the masculine space. Alex's production space, when accounting for the overlap between binary speakers, shows a clear analog to the proposed distribution of incongruence via ambiguity. I discuss Avery's productions more in Chapter 9 regarding multiple-signal productions, but their lack of incongruence via either method here holds throughout all variables examined.

Spotlight: Istus

Hypothesis **H1c**, regarding multiple signal dissonance, involves a much more complex interpretation of the vowel system as a whole, and I found no general support for it among nonbinary speakers in the productions of monophthongs. However, there are a few areas of interest, possibilities, and shortcomings of **H1c**, which are illustrated by Istus' monophthong production. All of Istus' vowels, not just the ones that differed significantly in the model results, can be seen in Figure 7.10, as the interpretation I provide here, such as it is, really benefits from this wider scope. Full visualisations for all vowels for all speakers can be found in Appendix F. The previous statistical analysis does not require such extensive visualisation, but taken as contextualisation for Figure 7.10 it is obvious that Istus is not producing prototypical vowels from either binary. However, it is unclear what strategy (or strategies) she *is* utilising. Her F2 productions of fronter vowels like FLEECE and KIT are quite high, while the F1 remains firmly in, or below, the norms for binary men. DRESS shows the opposite trend, consisting of a relatively masculine F2 and feminine F1. Many tokens of START, TRAP, and GOOSE exhibit extremely low F2, falling partially outside the norms of either gender. While there is not enough evidence to support a strong claim that Istus' monophthong productions can be interpreted as an example of multiple-signal incongruence, her data highlights some of the variegated complexity involved in the initial categorisation of signals on the part of the speaker, that is a prerequisite to being able to manipulate that signal at all. If this variation is due to gender, a number of possibilities emerge, of which I highlight only a few: Istus may

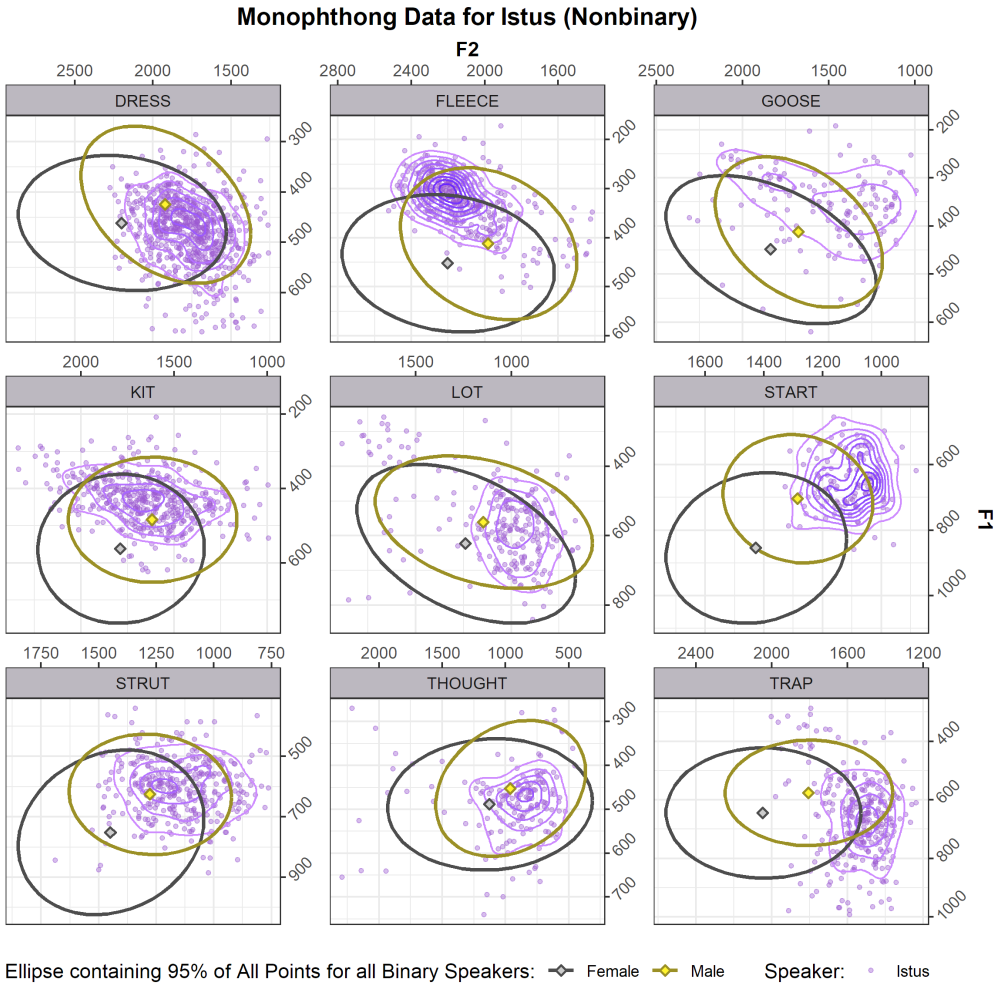


(i) Alex's Vowels plotted against all binary men and all binary women.



(ii) Avery's Vowels plotted against all binary men and all binary women.

FIGURE 7.9: Distributions of nonbinary speaker monophthongs compared to the data from binary men and women. Alex's (7.9i, top) production space is centralised relative to binary speakers, indicating incongruence via ambiguity. Conversely, Avery (7.9ii, bottom) shows no evidence of incongruent production of either type, as their vowels are produced within the binary male norms in all cases.



N = 683 (DRESS), 374 (FLEECE), 142 (GOOSE), 443 (KIT), 228 (LOT), 122 (START), 346 (STRUT), 164 (THOUGHT), 297 (TRAP).

FIGURE 7.10: Distribution of monophthongs from Istus, a nonbinary participant, plotted alongside data ellipses representing 95% of all tokens for men and women. Contour lines are set to display 10 "bins" - approximately 10% of tokens are within each contour. All monophthongs are shown, though only four (KIT, LOT, STRUT, and THOUGHT) show any statistically significant differences ($p < 0.04$) in F2 across gender. I present Istus' full vowels here (Full Vowels for all participants can be found in Appendix F) to illustrate the complexity involved in addressing **H1c**. Istus' vowels do not pattern cohesively with men, women, or any other nonbinary participants.

be signalling her gender differently for different vowels, or assigning different vowels varying weights in terms of the gendered signals or degree of gendering they convey. Within that, she may be interpreting (correctly, according to the literature) that one formant, F2, is more important than the other for conveying gender, and adjusting the amount of production effort (however you measure it) towards where it represents the most efficient payoff: If her F2 is within female norms, and F2 is a stronger cue for interpreting femininity, it may not matter so much exactly where her F1 falls. Regardless, her level of control of her vowels is in line with her other productions (including pitch, outlined in the previous Chapter), lending further support to **H3** and the influence of the individual condition.

7.2.5 Monophthongs: Discussion

Previous literature examining gender from a binary perspective (as I discuss in Section 7.2.2) on unnormalised data tends to describe F2 as being a stronger marker of physiological sex/gender difference than F1 (Weirich and Simpson, 2013). It is indeed the case that, examining vowels in isolation, that F2 shows significant differences across speakers while F1 does not.

The goal of normalising procedures in linguistic analysis is to mimic the perceptual process that extracts the critical features of vowel quality from the acoustic input, despite variation introduced by individual differences. In this way, it seeks to separate the physiological from the social, with the idea that if you are able to abstract productions away from the former, the only variation left is due to the latter, though see (Barreda, 2020) for a discussion of perceptual normalization as an active process by listeners that integrates these together to achieve perceptual constancy (p.2). I also problemetised this presumption of a clean separation in my review of the literature in Section 2.2.2, as per Johnson (2018a)'s assertion that even the most perfect normalisation technique is subject to researcher bias in its implementation. I asserted that applying normalisation techniques under a research paradigm that assumes a complete differentiation between the physiological and the social is going to result in the miscategorisation of a large amount of nonbinary speakers (as well as a subset of binary cis and trans speakers), and the erasure of a large amount of socially-significant information.

The results of the monophthong analysis support this theoretical point with concrete data. Normalisation is intended to erase presumably-static, presumably-asocial information from production, so that what remains is socially driven. However, the individual data presented in the spotlights using normalised data show that individual nonbinary speakers are utilising different production strategies to their binary peers. The quantile plot of the F2 of all speakers in Figure 7.6 illustrates clearly that, at least in this limited speaker data, nonbinary speakers are not just falling into a bimodal production space driven by their Assigned Sex at Birth (ASAB)[7], with the direction of significance falling out of "more nonbinary speakers patterning with binary male productions." That is, there is evidence in this analysis of *socially-motivated* manipulation of the overall F2, the very thing that normalisation erases as being non-malleable and a product of physiology (e.g. sex differences in vocal tract length).

A preponderance of fronted tokens that appear in the LOT vowel was found for many speakers. This was determined to be an artefact introduced by the production of the lexical item *cos* (a short form of *because*) being produced with something closer

[7]More nonbinary participants are Assigned Female at Birth (AFAB) than Assigned Male at Birth (AMAB), both in the vowel data and in the research overall.

to a SCHWA or STRUT vowel[8]. Removing *cos* from the dataset does not considerably change the outcome of the model reported in Table 7.2 – There are still no significant differences in F1, though F2 shows a change in p-value from 0.0209 (with *cos*) to 0.0359 (with *cos* removed).

Another point that emerges from this analysis of monophthongs is that, even on raw values, nonbinary speakers' production of F1 does not differ significantly from those of men or women. Even with nonbinary speakers removed, only 6/10 vowels (START, TRAP, KIT, LOT, FOOT, and STRUT) show significant differences in F1 between men and women, though removing the data of nonbinary speakers reduces the number of data points in this dataset by more than half, and exacerbates the differences in token numbers between individual participants in the dataset. This indicates the effect of (binary) gender on F1 production is relatively subtle. When examining individual speaker production across contexts, however, F1 varies more than F2 in virtually all cases. This may be partially explained due to the nature of the ItW recordings varying in the environmental loudness. Changes in the ambient loudness subsequently cause changes in speech, known as the Lombard Effect, and these changes are known to effect F1 production more than F2 production (Garnier and Henrich, 2014). However, while some speakers with high variation in F1 indeed have a high variation in the ambient noise of their recordings (such as Ollie, with about half of his recordings in a vehicle, was found to vary F1 significantly 39% of the time), this is not the case for all speakers with high F1 variation (such as Taylor, who recorded 5/6 of his ItWs in his bedroom, but whose variation rate is 30%).

Within the examination of monophthongs in isolation, I found no evidence among most nonbinary speakers of a compartmentalising of the vowel system in the way hypothesised to make multiple signal dissonance (**H1c**) possible. An examination of Istus' productions provided some initial evidence that multiple vowels can be manipulated together (and are therefore a potential site of incongruence via dissonance, as per **H1c**). However, further examination showed each of the vowels within the system were not conforming to the binary norms predicted by multiple-signal dissonance. Istus' productions are, therefore, more likely the result of a system-wide shift affecting multiple vowels, rather than individual manipulations. I leave the majority of the discussion of **H1c**, however, to Chapter 9, where I examine all analyzed signals together.

Overall, regarding variation across contexts and **H2**, when examined together, nonbinary speakers tend to exhibit more variation in monophthong production than binary speakers. However, there is much wider variation within groups than across them. In fact, the speaker with the least variation across recordings is nonbinary (Istus, whose excellent control over her productions I discuss in the previous chapter in Section 6.5.5), while the speaker with the most is a binary man (Jordon). Hypothesis **2**, therefore, is not supported strongly, and this data indicates that gender is a much weaker predictor of individual variation than originally proposed.

7.2.6 Monophthongs: Conclusion

Hypotheses **H1b** and **H1c** cannot be strongly confirmed nor refuted due to the lack of statistically robust differences between men and women, once the vowel data is normalised. The hypotheses being tested do not necessarily require gendered differences to be above the level of awareness, or indeed perceivable at all by participants, but they *do* require them to exist. For a signal to be utilised in the creation of gender

[8]This should have been removed with other stopwords in the initial cleaning, but many iterations of it remained due to its nonconventional/variable spelling(s); coz, cuz, 'cause, etc.

incongruence, that signal itself must be associated with gendered speech, either via direct utilisation or in carrying a salient stereotype of being used in gendered speech. Examining the unnormalised data reveals some evidence of difference across gender groups: women and men maintain different F2 productions that are consistent within-gender, with women's productions being higher than men's in all cases except FOOT. Closer inspection of individual data between different nonbinary speakers provides some support for the use of incongruence predicted by **H1b**. I have identified some features indicating the creation of gender incongruence via ambiguity (as in the case of Alex) and dissonance (as in the case of Kit). Further, while individual data from speakers like Istus seem to refute multiple-signal strategies as hypothesised in **H1c**, they do serve to illustrate some of the compounding complexities involved in the interpretation of multiple-signal presentations of identity, a theme which I return to in Chapter 9. The data shows only weak support for **H2**: While nonbinary speakers do exhibit more variation across contexts and interlocutors than binary men and binary women, the range of possible variation is much wider between individuals who share gender groups than across them.

At the top of the chapter, it was predicted that the widest variation in binary productions, and therefore the most likely areas for nonbinary speakers to manifest incongruence generally would be in F2 over F1, and in peripheral/extremely articulated vowels over more central ones, with both of these predictions pointing towards the most likely site of incongruence to manifest to be in the F2 of FLEECE. Gender has been implicated in sound change in NZE in multiple studies in the past, but this dataset was unable to confirm those previous results, likely due to a combination of the small number of speakers sampled, the relative subtlety of the gender effect, and the group-level demographic influence of age, as most of the speakers in RAINBO being essentially a generation younger than the youngest speakers in past NZE corpora. While previous findings are confirmed in that F2 productions were more strongly differentiated across binary gender than F1, this did not clearly manifest as a site for incongruent production. The four vowels found to show the most gendered differentiation are not considered to be the most peripheral – though arguably the *least* peripheral vowel, KIT, is implicated in NZE's push-chain shift, confirming previous literature and supporting the prediction of **H3**. The specific implication of the F2 of FLEECE was not supported, though it is intriguing that the nonbinary speakers' production of FLEECE, being significantly higher than the binary men's productions, bucked the trend: In all other vowels that are likely to show gender effects, nonbinary productions of F2 were significantly lower than the binary women's productions without distinguishing themselves from the male productions.

The most intriguing result of this analysis is the evidence that nonbinary speakers are manipulating their overall production space, indicating that subtle shifts in individual production may be less important than overall F2. It may be the case that a speaker can enact a relatively minor adjustment to the way they produce F2 overall, which results in much larger acoustic/perceptual results for less effort, as we see with Istus. Even if these results are not clearly aligned with gendered productions in the same way that subtle manipulation of single vowels along known gendered sound changes are, they are still certainly aligned with *intentional* productions. This is an exciting space for future work in sociolinguistics in incorporating nonbinary gender in production, and the evidence presented here is a clear challenge to the traditional dichotomy that frames the "psychological" and the "social" as two distinct parts that can be analysed separately.

7.3 Vowel Space Area (VSA)

7.3.1 Introduction

This section focusses on gender differences in VSA. In section 7.3.3, I outline the way incongruence is predicted to manifest among nonbinary speakers in terms of differences in the size and shape of their individual VSA and Vowel Space Conformity (VSC) when compared to those of men and women. The next section, 7.3.4, presents the statistical analysis and visualisations, while section 7.3.5 presents the results of those analyses and discusses their implications as well as some extensions and shortcomings of the study. In section 7.3.6, I provide some closing remarks.

7.3.2 VSA: NZE

While individual vowels (as explored in the previous section) can be implicated in gendered differences in sound change and driven by social factors, the overall vowel space is assumed to be a product of physiology. There is an extensive literature examining the links between VSA and sex/gender at a general level (as I explored in Chapter 2, Section 2.2.3). To summarise: VSA is found to generally be larger in women relative to men, children relative to adults, and in speech perceived as more intelligible. VSA is not often examined at the level of dialect or accent, excepting some recent examinations in dialects of American English (Fox and Jacewicz, 2017; D'Onofrio, Pratt, and Van Hofwegen, 2019; Jacewicz, Fox, and Salmons, 2007). There are no previous findings to suggest that differences in VSA are expected within NZE specifically, beyond those found in wider speech such as across gender, age, or register.

7.3.3 VSA: Hypotheses

Introduction

As VSA is the product of the interaction between multiple vowels, it can be examined either as a single gradient entity, or as the combination of multiple signals. A single gradient signal, represented by overall size of the vowel space, is covered under the predictions made by **H1b**. VSA as the product of multiple gradient entities, as embodied by the overall shape and topography of the space, is covered under the predictions of **H1c**. As we have seen in the previous section, it is possible for F1 and F2 to be influenced (or manipulated) somewhat independently from each other in a speaker's vowel productions. This is worth keeping in mind, as it may be the case that this process works in the other direction, as well: Some of the "lower-level" single vowel manipulations observed in Istus' monophthongs appeared to be the result of a "higher-level" manipulation of the overall vowel F2. It is also possible that manipulations of the overall space in size (VSA) and/or shape (VSC) may be the consequence of single-vowel manipulations resulting in a system-wide effect.

Hypothesis 1

Hypothesis H1b – Gradient Signal Incongruence *Given a single gradient variable, nonbinary speakers can communicate identity by producing incongruence either via **ambiguity** (in the form of more centrally-distributed tokens compared to binary male and binary female speakers), or via **dissonance** (by producing tokens in a bimodal fashion, where a single speaker's productions are within both binary male and binary female ranges).*

Under **H1b**, it is predicted that nonbinary speakers' mean vowel space area will fall between that of men and women, regardless of whether incongruence is created via ambiguity or dissonance. Figure 7.11 presents an idealised ambiguous presentation, in which a nonbinary speaker's vowels fall exactly between those of men and women. For illustrative purposes, the values for men and women are taken from Yang (1990) and use American English vowels. An idealised dissonant presentation is illustrated in Figure 7.12. In this approach, a given speaker is essentially producing two vowel spaces concurrently: One aligned with prototypically-feminine values, and one aligned with prototypically-masculine values (or as near to those prototypical values as that speaker is able to produce).[9] I have not plotted the potential convex hulls created by this type of distribution to make understanding the bimodal nature of each individual vowel more clear: Were it plotted using mean values, it would be identical to the ones found in Figure 7.11, whereas using some maximum/extreme measure would ensure the VSA selected only the most periphery vowels in each distribution. I explain how I approximate the latter in the next section, 7.3.4, which results in a hull stretched towards the periphery in similar ways for all speakers.

Hypothesis H1c – Multiple Signal Dissonance – *Given multiple signals, nonbinary speakers can communicate identity by producing incongruence across signals.*

It is possible that nonbinary speakers may consider the vowel system not as a holistic system, but as the product of multiple signals. In such a conception, each vowel is considered a single signal that can be gendered according to binary norms when examined in isolation, but that are deliberately produced in such a way that the resulting output is a VSA combining expected productions of both men and women. Figure 7.13 illustrates one such possibility predicted by **H1c**. A nonbinary speaker may produce tokens of FLEECE and LOT with the same values as binary women, for example, but produce THOUGHT, TRAP, and GOOSE using prototypically-male values.

If I posit that speakers can treat each vowel (or indeed, as in the previous section, each formant within each vowel) within the system as independent, that allows for the possibility that a single speaker's vowel system may exhibit any or all hypothesised forms of incongruence, though incongruence via ambiguity makes different predictions than incongruence via dissonance. However, should this be the case, that particular speaker is not treating VSA holistically, or at least not in a way captured by the predictions I have outlined here, and is therefore beyond the scope of the predictions (though not beyond the scope of **H1c**, which I address across the entire set of measured variables in Chapter 9).

Hypotheses 2 and 3

Hypothesis 2 – *Immediate contextual factors have a greater influence on nonbinary speakers than binary speakers.*

H2 predicts that nonbinary speakers will produce greater variation across contexts than binary speakers. However, **H2** is not addressed in this section. I did attempt to test its predictions, but two major limitations in the dataset made even tentative observations impossible: Firstly, for virtually all participants, at least one ItW recording is missing one or more vowels, rendering it impossible to measure VSA in these recordings at all. Secondly, because the binary participants' extended interviews were not transcribed and force-aligned due to time constraints, they have no "baseline" VSA to measure their recordings against.

Hypothesis 3 – *Variation is bound by the individual condition.*

[9] Note that Figures 7.11 and 7.13 use identical data and make the same predictions as presented in Figure 7.3 in the previous section.

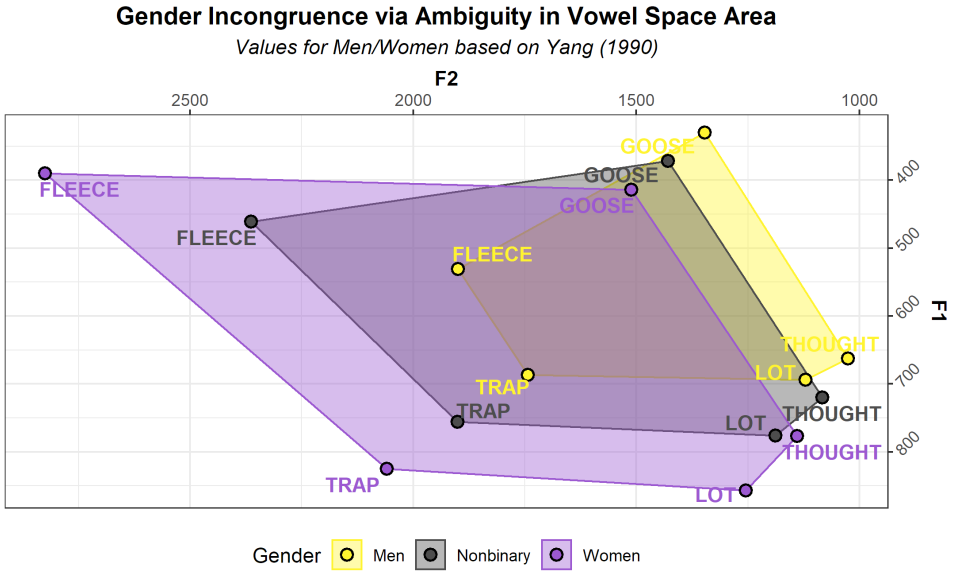


FIGURE 7.11: Simulated predictions of VSA variations across gender, given a nonbinary speaker producing incongruence via ambiguity. Values for men and women taken from Yang (1990), with nonbinary values extrapolated from there.

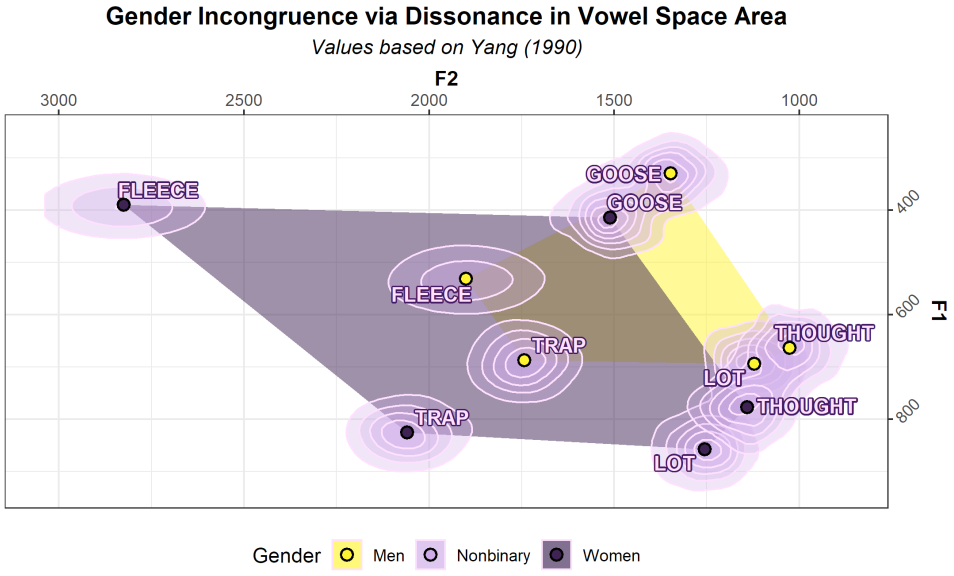


FIGURE 7.12: Simulated predictions of VSA variations across gender, given a nonbinary speaker producing incongruence via dissonance. The theoretical nonbinary speaker depicted is producing each vowel within the expected norms for both binary men and binary women. Values for men and women taken from Yang (1990), with nonbinary values extrapolated from there. Note that, were a convex hull drawn on this figure using the mean value, it would be in an identical position to the hull for nonbinary speakers pictured in 7.11 above, as the means of the idealised bimodal distribution would would fall between the two modes.

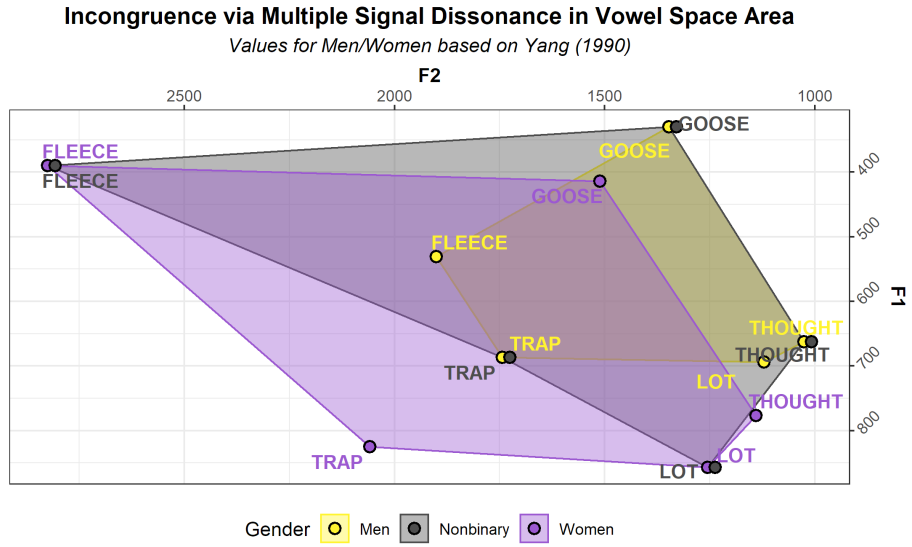


FIGURE 7.13: Simulated predictions of VSA variations across gender, given a nonbinary speaker creating incongruence via multiple-signal dissonance. Values for men and women are taken from Yang (1990), with nonbinary values extrapolated from there. Note this figure is identical to Figure 7.4 in the previous section, except with VSA emphasised via the colouring of the convex hulls, rather than monophthong production clouds.

H3 predicts that individual men and women's vowel spaces will be more similar to that of other participants that share their binary gender, relative to others, and that not all nonbinary speakers will create incongruence in similar ways. Should multiple nonbinary speakers produce a multiple-signal VSA (as in **H1c**), it is not required that each vowel be "gendered" the same way by each speaker.

7.3.4 VSA: Analysis

The same data and processing described in 7.2.2 was used here. Table 7.5 lists each participant's VSA, in Hz^2, using the mean location (F1 and F2) of each vowel. In addition to using this mean value, vowel space was measured a second time using a combination of the 25^{th} and/or 75^{th} quantiles of each value (see Figure 7.14). The mean of each vowel's F1 and F2 for each speaker were scaled around 0, and each vowel was assigned one of four quadrants based on the positive/negative status of the scaling. The motivation for this quantile measure was to obtain the maximal or relatively extreme measurement from each participant, emphasise variation across contexts in the direction of wider spaces, and potentially to distinguish the type of incongruence (if any) utilised by different participants, since taking the means only results in identical measures for both dissonance and ambiguity.

The vowel hull of incongruence via dissonance would look identical to the one utilising ambiguity if only the mean of each cardinal vowel were measured: That is, if a convex hull were drawn on Figure 7.12, it would be in an identical position to that found on Figure 7.11. In taking the quantile measurements, however, incongruence via dissonance in a given speaker may be identified due to it being much larger than the maximal VSA obtained from a speaker utilising incongruence via ambiguity.

Determining which of the two quantiles to use in each case was achieved using the following process: After outlier removal, I scaled each token based on the Hz

Gender	Name	VSA (Hz^2)
Women		
	Jesse	253 751.2
	Kelly	248 054.6
	Quinn	231 695.9
	Reese	126 456.5
	Remi	322 198.2
Group Mean		236 431.3
Men		
	Devon	105 324.6
	Jordon	152 588.0
	Noel	198 084.7
	Riley	114 226.0
	Taylor	199 817.7
Group Mean		154 008.2
Nonbinary		
	Alex	178 234.3
	Avery	159 902.1
	Istus	214 467.9
	Kit	309 732.3
	Mello	na
	Ollie	167 475.0
Group Mean		205 962.3

TABLE 7.5: Overall VSA in Hz^2 of all speakers, using the mean of the outermost vowels. Note Mello is not included due to a paucity of data.

values of F1 and F2 at the 50% (median) mark. This scaling applied a value to each token, with values closer to 0 representing closer proximity to the mean of the entire set of tokens. For each vowel, I then used the mean of both the scaled F1 and scaled F2 values to determine whether to take the 25^{th} quantile or the 75^{th} quantile: If the scaled mean was positive, I took the 75^{th} quantile measurement for that vowel, and if the scaled mean was negative, I took the 25^{th} quantile measurement. This divided the vowel space into four quadrants, with each vowel following one of four assignments based on which quadrant the scaled means appeared in. While the resulting shape is a quantile VSA, being composed of quantile measurements, I refer to it with the less accurate but more descriptive term "Maximal VSA", as which quantile it takes (25% vs 75%) is the one that would result in the largest total VSA.

After all measurements were obtained, a convex hull was created for each speaker from the F1 and F2 measurements from each vowel. This process selects an individual set of each speaker's outermost vowels, with vowels that would cause the hull to intersect itself discarded. This can be seen in Figure 7.15, which shows the mean and maximal VSA hulls calculated for three of the participants, as well as illustrating that different vowels may be selected for each speaker under this method. These points were then used to calculate each speaker's mean and maximal VSA in both Hz^2 and Bark squared ($Bark^2$). F1 and F2 Hz values in both datasets were converted to the Bark scale using the emuR package in R (Winkelmann et al., 2021), to better represent the perceptual contrast between vowel spaces. The Bark scale converts absolute pitch values between 20-15,500Hz to a value between 1 and 24, each value of which is

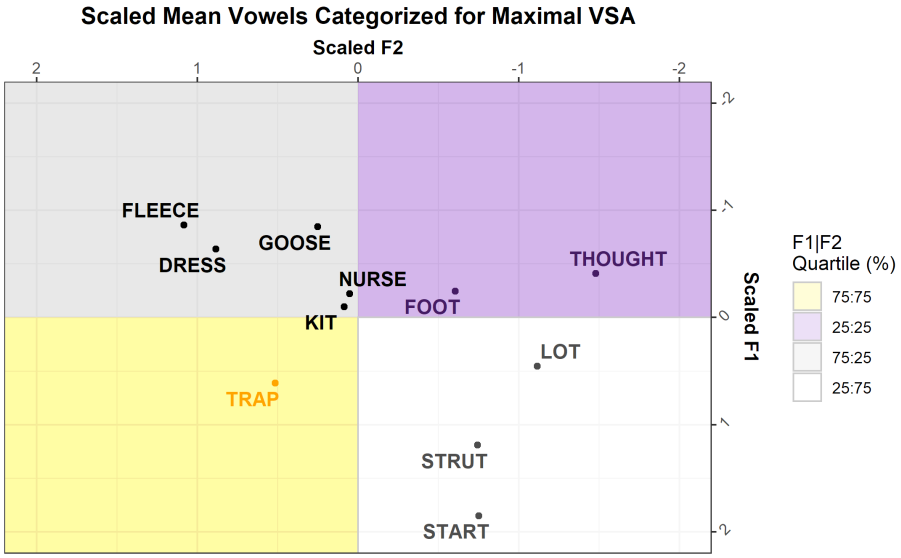


FIGURE 7.14: The overall VSA divided into quadrants, with the scaled means of each vowel per speaker plotted. The maximal VSA measurements for F1 and F2 were taken based on which quadrant each vowel fell into: Vowels in the upper right quadrant had the 25th percentile of both F1 and F2 taken, vowels in the upper left quadrant had the 25th percentile of F1 and the 75th percentile of F2 taken, etc.

equidistant perceptually (Zwicker, 1961; Weirich and Simpson, 2013). In addition to using "raw" Hz and bark values, I also converted both sets of data to euclidean distances. To obtain the euclidean distance, a mean was calculated for each vowel for each speaker[10], and the mean of all those points (the centroid) per speaker was also determined. The euclidean distance (d) between a given speaker's vowel mean ($F1_V$ and $F2_V$) and the centroid ($F1_C$ and $F2_C$) is calculated using the following formula:

$$d(V,C) = \sqrt{(F1_V - F1_C)^2 + (F2_V - F2_C)^2}$$

The mean of the distance between each vowel and the speaker's centroid was then taken as an overall measure of each speaker's VSA. This measure differs from the convex hull measurement in that it incorporates every vowel for each speaker into its calculation, while the convex hull uses only each speaker's most peripheral vowels.

Ultimately, the same data was shaped into eight different sets onto which the same simple linear model structure was fit to determine the relationship between gender and vowel space[11], reducing the statistical power of any results. Using the stats package in R (R Core Team, 2020), the model was run with VSA as the dependant variable and gender as the independent variable. Speech rate was also added as an independent variable, as it has been found to correlate negatively to VSA – that is, speakers with faster rates of speech tend to exhibit smaller VSAs (Fletcher et al., 2015; Tsao, Weismer, and Iqbal, 2006)

Manipulating the data in this many different ways, between Hz and the Bark scale , and area vs euclidean distance, warrants some comment. Fitting the models to both Hz and the Bark scale serves the purpose of attempting to distinguish productive

[10]Standard deviations for each of these were also measured to identify any obvious issues but were well within expected ranges, as outliers were removed at an earlier step.

[11]*lm([Area/Dispersion] ~ [gender] + [Speech Rate])*

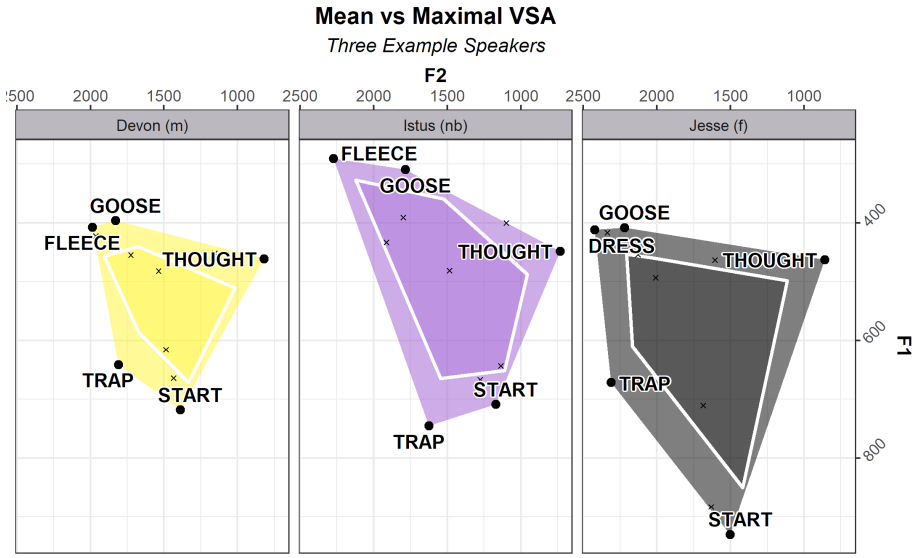


FIGURE 7.15: Comparing the two measures of vowel space across three example speakers. The inner shapes represent the mean VSA measurements per vowel, while the outer shapes represent the maximal VSA means. Named points show the vowels making up the outer edge of the vowel space, while x points show other vowels within the (expanded) VSA that do not contribute to the measurement (as they fall inside the area covered by the named points). Note that this results in the potential for different vowels to be used in calculating the convex hull across participants and mean/maximal measures.

difference (as Hz) from perceptual differences (as the Bark scale). While significant differences in models using the Bark scale would naturally imply differences in raw Hz measurements, significant differences in Hz are still informative as a measure of productive effort, even if they do not necessarily produce a perceptual result on the part of the listener.

Secondly, the division between using the hull-based calculation versus the euclidean distance is to examine differences in vowel space from the point of view of both overall area, as well as the dispersion of vowels *within* that area. When examining the vowel space using hulls, we are examining the relative distance between peripheral vowels, and getting a measure of vowel space size based on the extremes of that space. Euclidean distance also measures vowel space size, but from the point of view of the distance of each vowel from the center of the vowel space independently. Euclidean distance calculates a centroid for each speaker, and vowels are measured based on their distance from that shared central point, establishing an abstracted, but standardised, measure of vowel dispersion that is agnostic to shape. The benefits of exploring a model using data converted to euclidean distance is that it weighs the relationship between each vowel in the system and its distance from the centroid equally, regardless of that vowel's absolute position in the overall vowel space, or the number of tokens in each vowel. To illustrate, a speaker with 2 or 3 extremely peripheral vowels and the other vowels in the system clustered near the center may have a large VSA but small euclidean distance measure. Conversely, another speaker may not have extremely peripheral vowels relative to the other vowels in their system, but if all of those vowels are quite far from the centroid, that speaker would have a low VSA and high vowel dispersion. Ultimately, both area and dispersion measure the size of the vowel space, but one examines the absolute area while the other explores

the utilisation of the space within those boundaries. This allows a distinction to be made between two ways of measuring vowel space size: overall area (determined by the most extreme vowels) and overall dispersion (the mean distance between vowels from the center). A third way to examine an individual's vowel space is in terms of its absolute position in abstract space relative to other individuals or expected norms. While I present some visual analysis of Vowel Space Conformity (VSC) in the following section, I leave aside a more formal attempt at its quantification to Chapter 9.

However, it should be noted that this analysis varies from the established literature using euclidean distance in that I have not normalised the data at all across speakers. Generally, one of the purposes of euclidean distance in phonetics starts with normalised data so the resulting distance can be compared directly across speakers (Barreda, 2021): An equivalent distance between two points by two speakers represents the same amount of acoustic difference. Since the normalisation process has reduced the unwanted variation presumed to be the result of physiology, what variation remains, if any, must be the result of a socially-driven factor. As I have determined in the last section that there is good reason to suspect the variation within unnormalised data is socially meaningful, I have avoided normalising here.

7.3.5 VSA: Results & Discussion

The results of fitting these multiple linear models can be seen in Table 7.6. Only marginally significant effects appeared in the model predicting VSA using euclidean distances measured in Hz. When these measures are considered, binary men show a smaller overall VSA, and faster rates of speech result in smaller VSA as well. All other models show no significant effects when compared via ANOVA to a model with the gender term removed. I also reran the same models with nonbinary speakers removed. When this is the case, all significant effects of speech rate disappear, and men show a marginally smaller VSA ($p = 0.0457$, $R^2 = 0.426$) VSA than women in the model using the maximal quantile measures, with hull areas, in Hz. All other models show no significant effects. This is not unexpected, given the models are testing whether there are group-wide differences in vowel space area and vowel dispersion. They simply tell us that, with the exception of men (and speakers with higher speech rates) having less dispersed vowels overall, differences between speakers in this sample are not at the group/gender level.

Given the multiplicity of models run, and the marginal effect size, VSA is not a particularly robust indicator of gender among these participants. This is the case even when allowing for both productive and perceptual differences, normalisation, and restricting analysis to "known" binary populations. The weak statistical power of these models is a likely contributing factor, as each contains only one observation per speaker, for a total of 16 datapoints.

Figure 7.16 illustrates VSC via plotting each participant's vowel space against the average vowel space of all binary speakers, separated by gender category. The vowel spaces shown here use the measurements in Hz rather than the Bark scale, since it was the model using these values that showed significant, if minimal, dispersion effects. Visualising the data in this way makes a few things clear that are not apparent from the model results alone. First, that there is observable variation across all individuals within group: Reese, a binary woman, has a VSA similar in shape and much closer in area to Riley, a binary man, than to any other participant[12]. Kit, a nonbinary

[12]Reese: 126,456.5Hz^2; Riley: 114,226.0Hz^2; a difference of about 12kHz^2. For comparison, the next closest binary female, Quinn, has a VSA of 231,695.9Hz^2, a difference of over 100kHz^2.

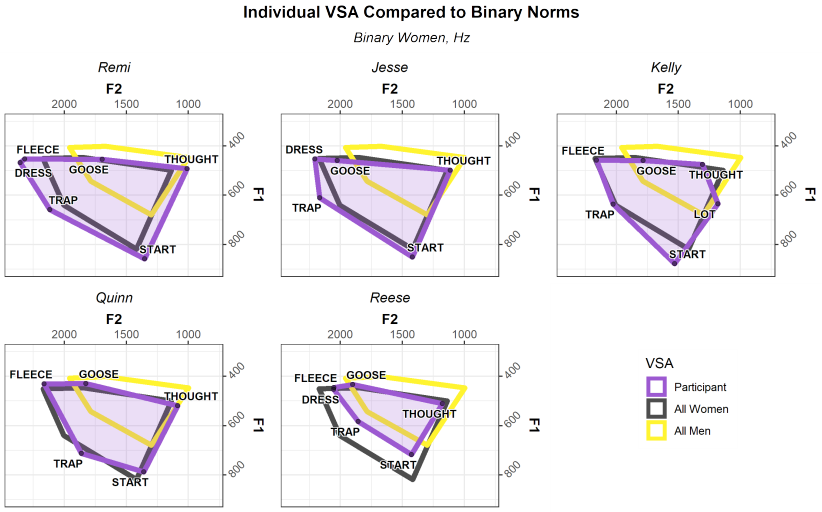
	Linear Models Run Using Hull Area Calculations			
	Mean VSA (Hz^2)	Mean VSA ($Bark^2$)	Maximal VSA (Hz^2)	Maximal VSA ($Bark^2$)
Intercept	631,469.000	20.792	926,187.700	28.188
(Nonbinary)	(229,951.500)	(7.395)	(363,479.800)	(11.154)
	p = 0.020*	p = 0.017*	p = 0.028*	p = 0.029*
Binary	5,646.082	−0.581	49,047.350	0.423
Women	(37,018.520)	(1.191)	(58,514.440)	(1.796)
	p = 0.882	p = 0.635	p = 0.420	p = 0.819
Binary	−63,807.790	−1.915	−90,204.030	−2.514
Men	(35,114.360)	(1.129)	(55,504.570)	(1.703)
	p = 0.097	p = 0.119	p = 0.133	p = 0.168
Speech Rate	−91,640.660	−2.849	−122,422.600	−3.206
	(49,244.280)	(1.584)	(77,839.450)	(2.389)
	p = 0.090	p = 0.100	p = 0.145	p = 0.207
Observations	15	15	15	15
R^2	0.458	0.360	0.493	0.352
Adjusted R^2	0.310	0.185	0.355	0.175
R.S.E. (df = 11)	54,599.570	1.756	86,304.440	2.648
F Statistic (df = 3; 11)	3.096	2.060	3.570	1.993

	Linear Models Run Using Euclidean Distances			
	Mean Distance (Hz)	Mean Distance (Bark)	Maximal Distance (Hz)	Maximal Distance (Bark)
Intercept	646.380	3.023	813.924	3.644
(Nonbinary)	(147.031)	(0.696)	(199.873)	(0.893)
	p = 0.002*	p = 0.002*	p = 0.002*	p = 0.002*
Binary	−17.696	−0.147	8.933	−0.045
Women	(23.670)	(0.112)	(32.176)	(0.144)
	p = 0.471	p = 0.217	p = 0.787	p = 0.763
Binary	−50.547	−0.223	−45.517	−0.184
Men	(22.452)	(0.106)	(30.521)	(0.136)
	p = 0.046*	p = 0.061	p = 0.164	p = 0.205
Speech Rate	−69.465	−0.304	−75.275	−0.289
	(31.487)	(0.149)	(42.803)	(0.191)
	p = 0.050*	p = 0.066	p = 0.107	p = 0.159
Observations	15	15	15	15
R^2	0.469	0.400	0.415	0.282
Adjusted R^2	0.324	0.236	0.256	0.087
R.S.E. (df = 11)	34.911	0.165	47.458	0.212
F Statistic (df = 3; 11)	3.241	2.445	2.604	1.442

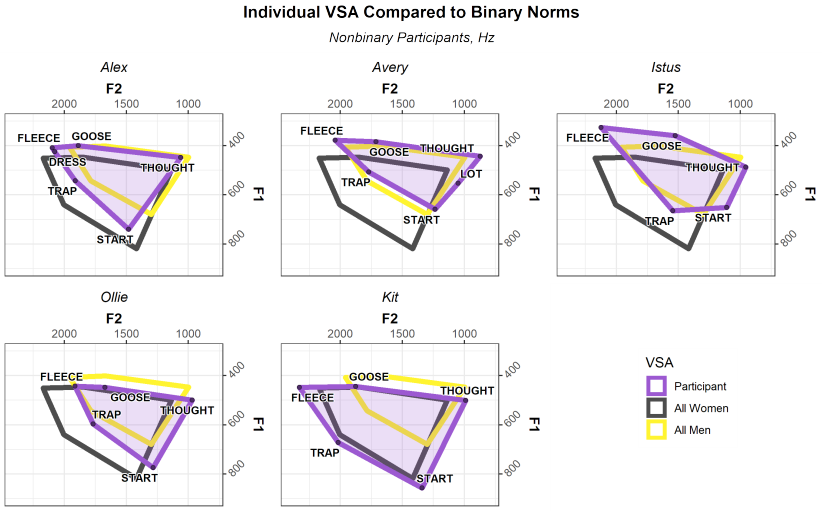
TABLE 7.6: VSA Linear Model Results. As these are models run on measures of total vowel area and dispersion, there is only a single observation per speaker. Expectations as to the meaning of significant differences should be adjusted accordingly.

Table generated with the Stargazer R package (Hlavac, 2018).

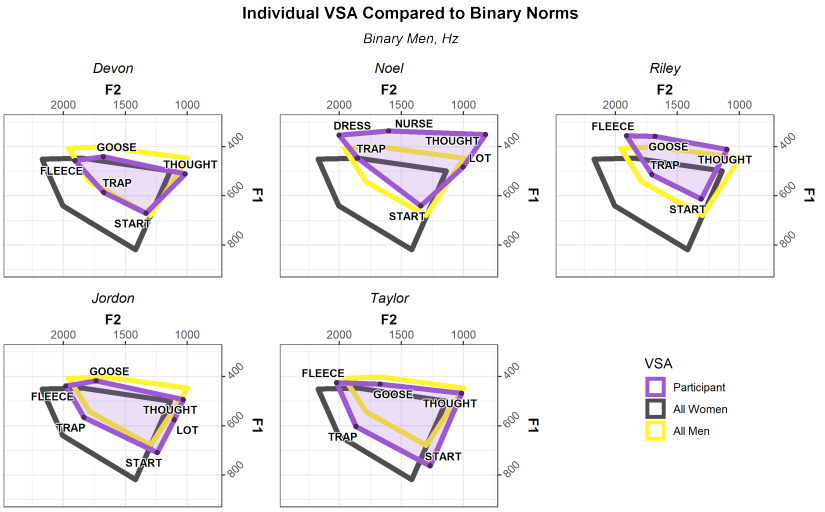
FIGURE 7.16: Visualising Vowel Space Area per participant, categorised by gender, compared to averages of all binary participants, using Hz values.



(i) Binary Women.



(ii) Nonbinary Participants.



(iii) Binary men.

participant, has the second largest VSA overall, and firmly encompasses the entire female average while avoiding encroaching into the extremely lowest F1 and F2 combinations of the "exclusive" male space. Alex (and to an extent Ollie), on the other hand, present a VSA consistent with a somewhat ambiguous presentation, as their VSA is mostly within the area of overlap for both men and women. Note also the similarity between Ollie and Taylor's distributions. Istus presents something new entirely, displaying extreme productions (most notable in the F1 of TRAP and the F2 of FLEECE and START) that create a shape unlike other participants, even other nonbinary participants. In this way, **H3** is supported from the point of view of nonbinary speakers utilising different strategies from each other to present identity, but also introduces some confounding evidence that binary speakers are also utilising different strategies from others in their cohort in using VSA to present identity.

7.3.6 VSA Conclusion

It is clear from the data that the VSA of many nonbinary speakers does not align with norms for either binary sex/gender. It is also clear, however, that the VSA of binary people does not always neatly align to binary norms, either. Were the size and shape of the vowel space a distinct product of physiology, it would be expected that speakers, regardless of gender, would fall neatly into two categories of "large VSA" and "small VSA", or "masculine dispersion" or "feminine dispersion" based on ASAB. It is evident from the data that they do not, and in fact the binary men and women do not appear to be so constrained, either. Additionally, in terms of overall size, wide variation is found both within and across gender groups. In terms of dispersion, binary gendered speakers do tend to align more closely with binary means, partially as a product of the size of the data set meaning each binary speaker contributes substantially to the creation of those means.

Unfortunately, the complexity and paucity of the data make it impossible to test *any* of the three hypotheses using robust quantitative methods, though I have presented some interesting trends (and possible directions for future work). While visualising the VSC (as in Figure 7.16), it seems clear that some nonbinary participants are producing a vowel spaces unlike the patterns found in men and women, and there is unexpected variation within binary gender groups as well. I believe this may be an indication of overall vowel space topography using (or being indicative of the use of) multiple signals in presenting gender. I return to this point of interest in Chapter 9, where I discuss Hypothesis **H1c** when examining dissonance across multiple signals.

7.4 Spotlight: NEAR/SQUARE

7.4.1 Introduction

This Spotlight should be taken as a kind of "side-quest", a brief deviation from the main hypotheses to highlight some interesting data from the older triad of speakers, Alex, Kelly, and Noel. These speakers are of an age where they are predicted to be taking part in a gendered vowel merger in NZE between the NEAR and SQUARE vowels, which has been studied extensively in the literature using binary gender categorisations. After backgrounding this vowel shift, in lieu of predictions regarding incongruence I instead attempt to outline some predictions based on the participants' situation in time. While the results and analysis of data from three speakers is

not enough to form conclusions, it does serve to correct a possible omission in the historical record based on "new" ways of categorising participants.

7.4.2 Background

The merging of the NEAR and SQUARE vowels in NZE is well represented in the literature, with multiple production and perception studies conducted over the last half-century. Gordon and Maclagan (2001) and Hay, Drager, and Warren (2009) offer an excellent overview of the literature regarding the merger in NZE, which I paraphrase only briefly here.

From at least the 1850s, some minority of New Zealanders merged the vowels found in the NEAR and SQUARE lexical set, producing words like *rare* and *rear* as homophones (Gordon and Trudgill, 1999). Multiple studies from around 1980 onwards, collecting recordings from participants of different ages, found evidence of a merger by approximation with both vowels being produced as NEAR. Women and working class speakers were found to be the most likely to merge, with a brief shift in the mid to late 1980s (representing speakers born around 1970) towards a merge on SQUARE (Holmes and Bell, 1992). Confounds in the analyses of this time, as pointed out by the original authors, include problems with determining class in NZ at the time (Gordon and Maclagan, 2001, p. 225), lack of data (Gordon and Maclagan, 2001, pp. 224–225), the variable reporting of the direction of the merge (Holmes and Bell, 1992), and some competing findings which found no gendered differences in production (Bayard, 1987; Gubian et al., 2019). Production of both vowels settled on NEAR via lexical diffusion (Maclagan and Gordon, 1996; Gubian et al., 2019) for most speakers by the 1990s (Gordon and Maclagan, 2001; Gubian et al., 2019, p. 222) and have presumably stayed there, though see Hay, Drager, and Warren (2009) for potentially competing evidence showing that the youngest speakers in the sample may be beginning to re-distinguish these vowels. In addition to correlational links with identity, the production of the NEAR/SQUARE merger in NZE has been found to be sensitive to contextual (possibly accommodation) effects: Hay, Drager, and Warren (2009) found that speakers were more likely to merge the vowels when the experimenter also merges those vowels, and maintain more distinction between them if the experimenter they meet does so as well.[13] Figure 7.17, reproduced from Bauer et al. (2007), shows the assumed "contemporary" (as of ~15 years ago) productions of these vowels.

The latest year of birth of the participants in the cited studies is 1982.[14] The younger participants in my project were born almost 20 years after the youngest speakers studied in other work, long after the point where there seems to be measurable gendered differences in NEAR/SQUARE production in the literature. However, the older triad of Noel, Kelly, and Alex, with years-of-birth from 1965-1975, are within the highly variable group reported by Gordon and Maclagan (2001) (born around

[13] While accomodation effects are somewhat controlled for in my own research (I have a distinction between NEAR and SQUARE vowels, and conducted all of the interviews myself within a few months of each other), I have not taken any steps to account for possible differences in the degree or direction of participant accommodation, both to me or to the varied interlocutors in their ItW recordings. In the latter case, it is impossible to know the speech features of the interlocutors that participants may be accommodating towards.

[14] Hay, Drager, and Warren (2009) does not report years of birth, but does contain many speakers in their 20s. Given the publication date of 2009, it is likely these speakers were born closer to 1990. However, the younger speakers in this study both tended to maintain a greater distinction between the vowels and were also influenced by the dialect of the experimenter.

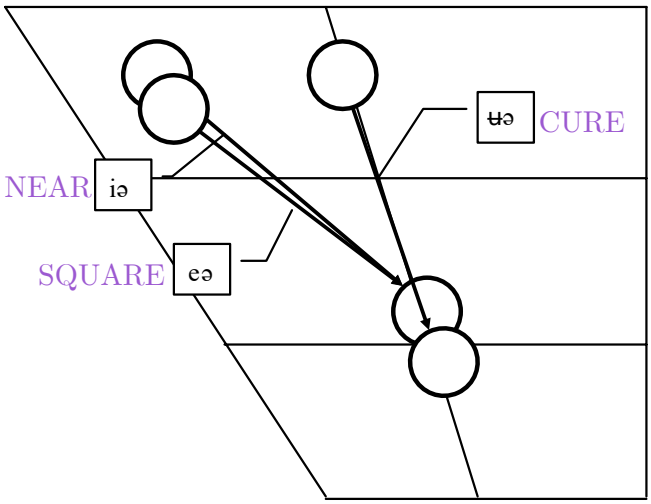


FIGURE 7.17: Approximate start/finish targets for centralising diphthongs in contemporary NZE. Reproduced with permission from Bauer et al. (2007): I have added NEAR/SQUARE/CURE lexical set labels to the IPA to make comparisons easier.

1970, recorded in 1983)[15]. In their studies, this group showed over half of participants producing variable vowels (a partial merge), 16% fully merged on either vowel, and 33% producing NEAR and SQUARE distinctly.

7.4.3 NEAR/SQUARE Predictions

Given the variability of the data from previous studies, and the small number of participants I examine here (3 individuals), it is inappropriate to make predictions regarding incongruence.

Single speakers should not be taken to stand for entire populations. Noel, Alex, and Kelly are themselves outliers within the research cohort in terms of age, and we have seen this manifest in other aspects of production. Noel in particular displays a vowel space that is dissimilar to the other men, and we shall see in the following chapter that both Kelly and Alex display relatively anomalous frication. That said, given the literature on speakers of similar years-of-birth examined above, and assuming some adherence to the apparent-time hypothesis (that older speakers use "older" variants of words, reflecting the language they learned in childhood (Bailey et al., 1991)), we may expect the following: all speakers are likely to produce some variability across tokens. Noel, the binary man, is likely to merge on NEAR; Kelly, the binary woman, is likely to produce more tokens of NEAR as SQUARE (that is, merging on SQUARE) as per Holmes and Bell (1992, p. 259). Variable predictions abound for Alex, the nonbinary participant.

1. They may speak in line with (or similar to) Kelly, who shares their ASAB, which we can interpret as an analog for their gender in the past as they understood it, indicating a strong influence of acquisition and therefore evidence for the apparent time hypothesis.

2. Their speech may utilise gender incongruence in relation to binary speakers born around 1970, which would also align with the apparent time hypothesis.

[15]Participants in my study self-reported their age at the time of data collection, but I did not collect their actual year/date of birth, and so I extrapolate it a little here.

3. They might also speak in a way that utilises gender incongruence, but in relation to *contemporary* binary productions, indicating a shift in speech happening much later in life.

Unfortunately, given we have no way to know Alex's "starting point", and that the merge is complete and consistent for all younger speakers, these questions are unanswerable given the current data. However, they represent a compelling direction for future research, of which this analysis serves as a small starting point.

7.4.4 NEAR/SQUARE Analysis

	Name	Gender	Vowel						Total	
			NEAR		SQUARE					
					Yeah		Other			
	Kelly	Female	77	(51)	338	(306)	106	(44)	521	(401)
	Noel	Male	77	(48)	177	(126)	49	(14)	303	(188)
	Alex	Nonbinary	198	(50)	959	(439)	183	(57)	1340	(946)
Total			352	(149)	1474	(1271)	338	(115)	2164	(1535)

TABLE 7.7: Number of tokens of NEAR and SQUARE per participant with F1 and F2 measures taken at the 50% point of production. The smaller numbers in parentheses represent tokens with complete measurements, of both F1 and F2 at 25, 50, and 75% duration. If a given token is missing a measure at 25 or 75%, or if the Hz recorded at either of those time points was later removed as an outlier, the token was also removed from the plot in Figure 7.19.

Tokens of NEAR and SQUARE produced by the older triad were extracted and cleaned following the same process outlined in Section 7.3.4, with one extension to account for the data being diphthongs rather than monophthongs: In addition to extracting the F1 and F2 at the half-way mark of each value, F1 and F2 was also measured at the 25% and 75% timepoint of each vowel. Tokens of *yeah* were also removed, as they were overrepresented. Formant measurements were normalised using the modified Lobanov method explained in Section 7.2.2. Table 7.7 lists the number of tokens for each vowel, for each participant.

To determine whether each participant merged NEAR and SQUARE, mixed effects linear regressions were run for each participant using the lme4 package in R (Bates et al., 2015). Formant frequency in Hz (F1 and F2) was treated as the dependent variable, vowel (NEAR/SQUARE) was the independent variable, and word was treated as a random effect. The results of these models are visualised atop raw Hz measurements in Figure 7.18, along with the plotted values for each speaker at each timepoint. Full model results can be found in Table 7.8.

7.4.5 NEAR/SQUARE Results & Discussion

The results of the modelling show that Noel has a complete merger on NEAR, as at no point are Noel's vowels distinguishable from each other. Both Kelly and Alex maintain a distinction, but the manifestation of that distinction is quite different across the two participants. The features of this difference becomes more apparent when examining the model results in conjunction with the raw vowel trajectories, as plotted in Figure 7.19. Alex's productions of both vowels have different starting and

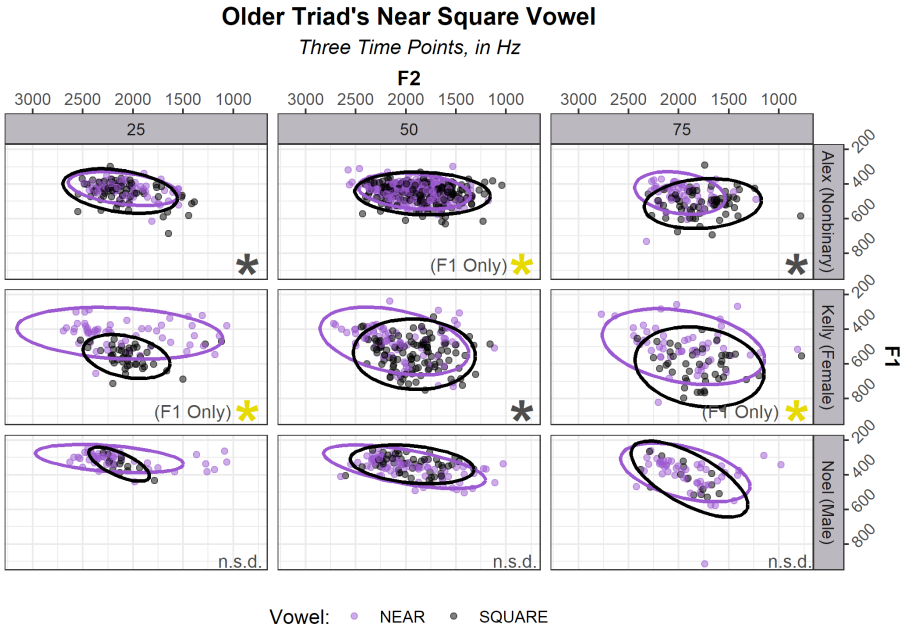


FIGURE 7.18: All three timepoints of NEAR/SQUARE plotted, alongside information from mixed effect linear regression models. Each facet represents two distinct models, with F1 and F2 as dependant variables and the same speaker, timepoint, and random effect of word. Facets with black asterisks indicate models where both F1 and F2 were significantly differentiated between NEAR and SQUARE. Yellow asterisks indicate models where only F1 differed significantly. Facets labelled n.s.d. (e.g. all of Noel's productions, along the bottom row) indicate no significant difference between the vowels at all time points.

ending points, but show only a marginal distinction ($p = .0551$) in F2 at the middle point, indicating different diphthong trajectories which overlap in the middle. Kelly maintains a statistically significant distinction between NEAR and SQUARE for both F1 and F2 in both vowels at the 50% timepoint, but only in vowel height (F1) at the other timepoints.

According to the distinction made in Gordon and Maclagan (2001, pp. 220–221), using the production of DRESS to track the direction, Noel is merging on NEAR and Kelly is maintaining a distinction between NEAR and SQUARE. Alex, however, would be considered to be variable or merging (in the process of merging) the diphthongs on SQUARE. While Noel's completed merge just indicates within-category variation, Alex's maintenance of the distinction between the vowels, while still producing wide variability (especially along F2), is evidence that Alex and Kelly belong to very different populations. What is unknown, however, is whether that population difference is related to their difference in gender and changes in speech over time, or whether Alex would have been classified in the past literature as a "variable" speaker.

I did run separate models comparing *yeah* to both NEAR and SQUARE separately, and found that, for the most part, all three speakers maintained an F1 distinction between *yeah* tokens and SQUARE vowels. Alex and Kelly additionally maintained a distinction between *yeah* tokens and NEAR vowels, while Noel produced *yeah* tokens as NEAR. This is likely due to the merger proceeding by lexical diffusion, as has been attested elsewhere (Maclagan and Gordon, 1996). The comparison of *yeah* with the participants' other NEAR and SQUARE vowels can be found in Figure 7.20.

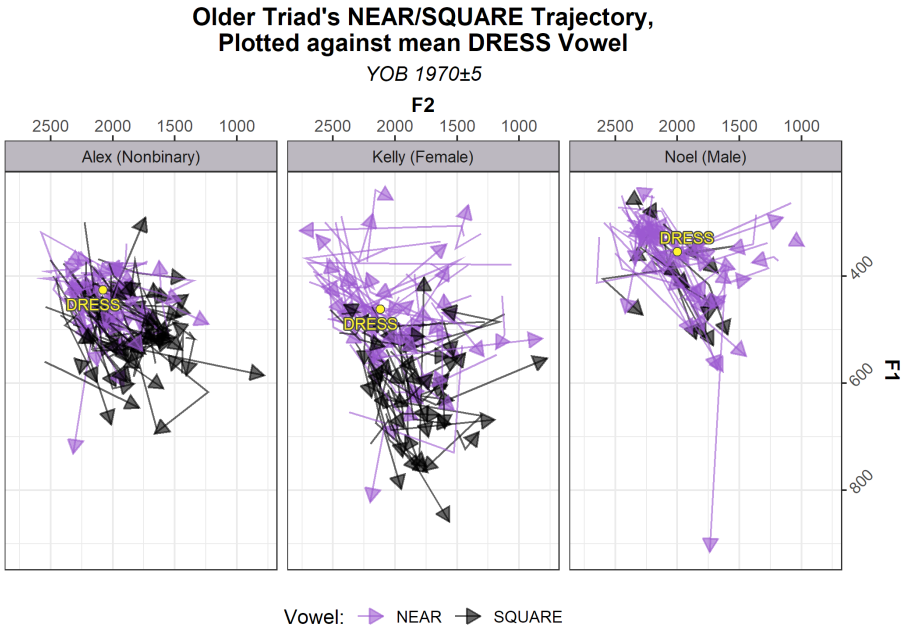


FIGURE 7.19: Vowel trajectory of NEAR/SQUARE vowels among the older triad with tokens of *yeah* removed, and the mean of each speaker's DRESS at the 50% timepoint plotted (yellow) to aid in illustrating the direction of change, as per Gordon and Maclagan (2001, pp. 220–221)

While the lack of evidence of past productions makes it impossible to draw direct conclusions about the influence of nonbinary gender on past language, the obvious differences across speakers in this triad are nevertheless fascinating. Alex's variable productions actually conform to the majority of speakers in the original study discussed in Gordon and Maclagan (2001), recorded in 1983. While we cannot know what Alex (or Kelly, or Noel) sounded like in 1985 without a time machine, some possibilities emerge: Alex was already producing these vowels as a (latent?) presentation of their nonbinary gender; Alex originally produced vowels closer to the norms expected from their ASAB, and shifted their vowels in response to a change in their identity. If the latter, the shift to a variable production may indicate an orientation to a "past" time point, rather than a contemporary one (since the evidence from younger speakers points to a completed merger for both genders). Whether these differences would have existed if these three speakers had been recorded as children or teenagers during the studies in Gordon and Maclagan (2001) is impossible to know. However, the current results offer potentially testable challenges for the apparent time hypothesis, or the concept of gender-based leaders and laggers in sound change differing based on the acquisition (or rejection) of gender.

7.4.6 NEAR/SQUARE Conclusion

This section established that there are significant differences between the production of the NEAR/SQUARE diphthongs among the older triad. The sample of speakers is small, and represents a volatile time in the trajectory of a sound change. One of the speakers is, as far as the literature is concerned, the *only* (or, at least, first) nonbinary speaker in this cohort, and there is no way to know how they sounded when they were younger. An impossible question using meager data to draw no conclusions –

		Model	AIC	BIC	Log Likelihood	Deviance	χ^2	Pr(>χ^2)	Significance
Alex	25%	F1	184.7772	196.0259	−88.388 61	176.7772	11.945 43	0.000 547 8	***
		F2	431.6145	442.8633	−211.8073	423.6145	4.758 679	0.029 150 9	**
	50%	F1	567.0406	582.8118	−279.5203	559.0406	6.773 155	0.009 253 9	***
		F2	1271.918	1287.689	−631.9590	1263.918	0.831 708 6	0.361 778 9	
	75%	F1	170.0752	181.1917	−81.037 62	162.0752	15.329 52	9.03e-05	***
		F2	427.0454	438.1619	−209.5227	419.0454	5.442 838	0.019 648 7	**
Kelly	25%	F1	186.5287	196.9091	−89.264 34	178.5287	20.907	4.8e-06	***
		F2	320.4850	330.8655	−156.2425	312.4850	2.210 865	0.137 041 8	
	50%	F1	383.9514	396.7894	−187.9757	375.9514	13.599 52	0.000 226 2	***
		F2	561.6043	574.4423	−276.8022	553.6043	4.287 896	0.038 384 6	**
	75%	F1	179.3004	189.5993	−85.650 21	171.3004	19.875 25	8.3e-06	***
		F2	313.9533	324.2522	−152.9767	305.9533	1.517 638	0.217 977 2	
Noel	25%	F1	118.0043	126.5128	−55.002 14	110.0043	0.718 870 1	0.396 514 8	
		F2	182.922	191.4305	−87.460 99	174.922	0.245 017 7	0.620 605 3	
	50%	F1	234.3819	245.7271	−113.1910	226.3819	1.468 501	0.225 582 4	
		F2	367.4068	378.7519	−179.7034	359.4068	0.033 285 6	0.855 234 8	
	75%	F1	119.7615	128.3970	−55.880 75	111.7615	0.664 427 4	0.415 001 2	
		F2	189.6197	198.2552	−90.809 85	181.6197	0.029 858 7	0.862 811 2	

TABLE 7.8: NEAR and SQUARE Model Fitting. Each model contains a random effect of word.

why examine it at all? I include this inconclusive examination of production within the older triad of speakers to attempt to produce, however small, some representation of nonbinary gender within the established literature on NZE.

7.5 Vowels: A Summary

I provide here a brief summary of the results of this chapter examining the gendered production of vowels.

The statistical analysis of the monophthongs found some weak support for variation by gender in the F2 of four vowels (THOUGHT, KIT, LOT, and STRUT). This is in contrast to the literature's suggestion that the largest differences in gender are found in the vowels representing the more extreme tongue and jaw positions, namely FLEECE, GOOSE, and TRAP. These vowels did show some evidence of variation, but their p-values were much closer to 0.05, and therefore their explanatory power is relatively weak compared to that found for THOUGHT, KIT, LOT, and STRUT. The examination of F2 by gender in unnormalised data found that women did differ significantly from both nonbinary speakers and men. However, even these minimal indications of difference across men and women disappear when the data is normalised using speaker-intrinsic methods before recategorising speakers by gender. Examinations of individual nonbinary speakers found evidence of incongruence via both ambiguity (in the case of Alex) and dissonance (in the case of Kit). Some complications arise in the data in that one nonbinary speaker, Istus, produces vowels unlike any other participant, highlighting some problems in the predictions made by **H1c** regarding multiple-signal dissonance but confirming those made by **H3** regarding the individual condition. The examinations of monopthongs did find minimal support for **H2** regarding nonbinary speakers' exhibition of greater variation across context, but this is mitigated by the fact that variation within every gender group was much wider than the variation found across any two gender groups.

The examination of VSA found it to be a weak indicator of binary gender, and therefore an unsuitable measure by which to compare nonbinary speech production.

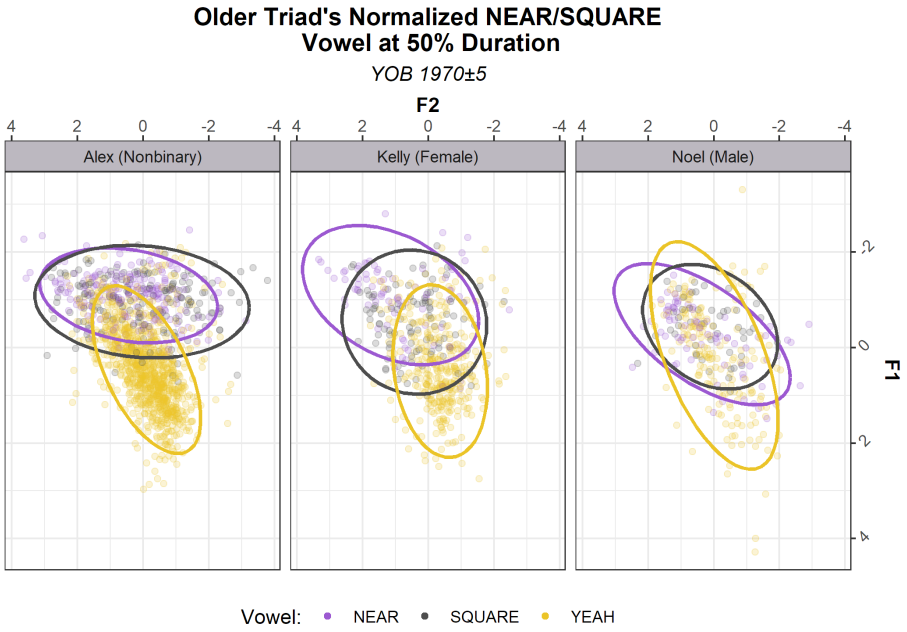


FIGURE 7.20: Normalised tokens of NEAR, SQUARE, and <*yeah*> at the 50% timepoint of F1 and F2 among the older triad. Ellipses are drawn around 95% of the tokens for each vowel/word, for each speaker.

However, in examining vowel dispersion via VSC it found that binary speaker productions conform more closely to binary norms than they do for VSA as measured by area alone.

Finally, the examination of NEAR/SQUARE was not intended to address the hypotheses so much as provide an example of ways in which the historical record on language variation in NZE can incorporate nonbinary speakers. However, this examination did provide some weak support for **H1a** in that Alex's near-merged productions of NEAR/SQUARE are consistent with incongruence via ambiguity, when examined in the context of the older speakers (though not in the context of the younger speakers, who have all merged these two completely). However, these productions are also consistent with a "variable" production attested in the past among both binary genders.

7.6 Vowels: Conclusion

In this chapter, I examined the effects on nonbinary gender on the production of vowels in NZE. I analyzed the production of eleven monophthongs, the VSA and VSC for every participant, and examined the production of the NEAR/SQUARE merger in the older triad of speakers. I found some indication that some nonbinary speakers utilise incongruence in presenting gender identity in monophthong production. I also found that, while contextual factors seem to have a greater effect on monophthong variation in nonbinary speakers, individual variation is considerable, regardless of speaker gender. Finally, I have provided a small addition to the historical record, in the form of an analysis of a recent merger in NZE to include, for the first time, the production of a nonbinary speaker. Having examined the creation of incongruence in finely gradient productions, I turn now to a more categorical variant implicated in gendered speech in NZE: intervocalic /t/ frication.

Chapter 8

Intervocalic /t/ Frication

8.1 Introduction

This chapter examines the use of frication in intervocalic /t/. Section 8.2 defines the variable of interest and provides an overview of the social distribution of frication in New Zealand and Australian English. Section 8.3 presents the predictions made by the relevant hypotheses regarding incongruence and production context. Section 8.4 details the way tokens were categorised, then gives an overview of their distribution across speakers in terms of descriptive differences, summary statistics, and more quantitative modelling of the data. I then explore a number of spotlights in Section 8.5, where I take a more qualitative approach in addressing the social messaging and intention of frication across individuals and generations of speakers. Section 8.6 returns to the hypotheses and addresses whether, and to what degree, the predictions were confirmed or refuted. I conclude with Section 8.7, summarising the chapter's contents and addressing some shortcomings and directions for future research.

8.2 Background

In New Zealand English (NZE), /t/ is a voiceless stop consonant, produced when the tongue forms a closure and subsequent release at the alveolar ridge. Word-initially, before a stressed vowel, /t/ is aspirated [t^h], produced with high frequency aperiodic noise between the stop closure and the onset of voicing of the following vowel. Intervocalic /t/ appearing within words in NZE can be produced with a variety of realisations, though the most frequent are the canonical unaspirated form, [t], and, more recently, the alveolar flap [ɾ]/[d] frequently found in other varieties of English. Other documented intervocalic productions include noisier variants like aspirated [t^h], pre-aspirated [^{h}t], and fricated/affricated forms, as well as the more lenited variants, glottalised and glottal stops [ʔ], and, sometimes, the stop is deleted completely ∅ (Hay and Foulkes, 2016).

Much of the work on intervocalic /t/ in NZE is concerned with changes in its production over time. From the historical canonical [t] assumed to be present at colonisation (Hay and Foulkes, 2016), the voiced variants [ɾ]/[d] first appeared among working class men, presumably due to the influence from Australian and/or Standard American English (Bell, 1977). From there, according to Bell (1977, 1984) and Holmes (1995), it spread to the rest of NZE. While [ɾ]/[d] is the most common variant found today (including in my own corpus), it is still correlated strongly with male speakers, working class speakers, and young speakers.

While flapping then began to spread from working-class men throughout the population, noisier/fricated variants (at the other extreme acoustically) became associated with middle- and upper-class women (Fiasson et al., 2016). It is unclear whether

the fricated forms indicate a true increase in noise in the production of intervocalic /t/ for some speakers, or whether this variation has always been present and it is simply the juxtaposition with growing rates of flapping among younger people that has caused a shift in the perception of this form as marked. Fricated and aspirated (including preaspirated) variants are assumed to carry some form of prestige, and are used more by older women, professional people, and gay men (Tollfree, 2001; Holmes, 1995; Taylor, 1996).

8.3 Hypotheses

As a variable, intervocalic /t/ can be treated as either gradient or categorical along a number of dimensions. The degree or simple presence/absence of voicing, stop closures, or high-frequency noise, as well as the relative location of each of those things all represent possible dimensions upon which social information may be attached. While I begin with a relatively gradient categorisation of /t/, I eventually collapse them to a binary category, and therefore address intervocalic /t/ frication as predicted by **H1a**.

8.3.1 Hypothesis H1: Acoustic Incongruence

Hypothesis H1a – Categorical Signal Incongruence

> *Given a single categorical variable, nonbinary speakers can communicate identity via **dissonance**, by producing tokens at rates between those found for binary male and binary female speakers, or via **ambiguity**, by producing a variant not found in either men's or women's productions.*

As outlined above, in the New Zealand English context, flapping intervocalic /t/ is associated with a constellation of traits that includes masculinity, youth, and being working class. Highly fricated /t/, however, including preaspiration, is associated with a constellation of features that encode a particular form of older, high-status femininity.

Gender incongruence via ambiguity is created by the use of signals that are equally applicable to either binary norm. With gradient signals, this is easily conceptualised as encompassing the middle ground between the two gendered distributions. However, in regards to categorical signals, equal *inapplicability* between binary norms becomes another relevant ground for creating incongruence via ambiguity.

H1a predicts that nonbinary speakers seeking to create incongruence via dissonance will produce more flapping than women and less flapping than men, and that nonbinary speakers seeking to create incongruence via ambiguity will produce a majority of forms that are neither fricated nor flapped. These possible outcomes are visualized in Figure 8.1, reproduced from Figure 3.1.

8.3.2 Hypotheses 2 and 3

Hypothesis H2 – Contextual Influence

> *Immediate contextual factors have a greater influence on nonbinary speakers than binary speakers.*

H2 predicts that individual nonbinary speakers will exhibit more variation across contexts, compared to individual binary speakers. While I discuss the particulars

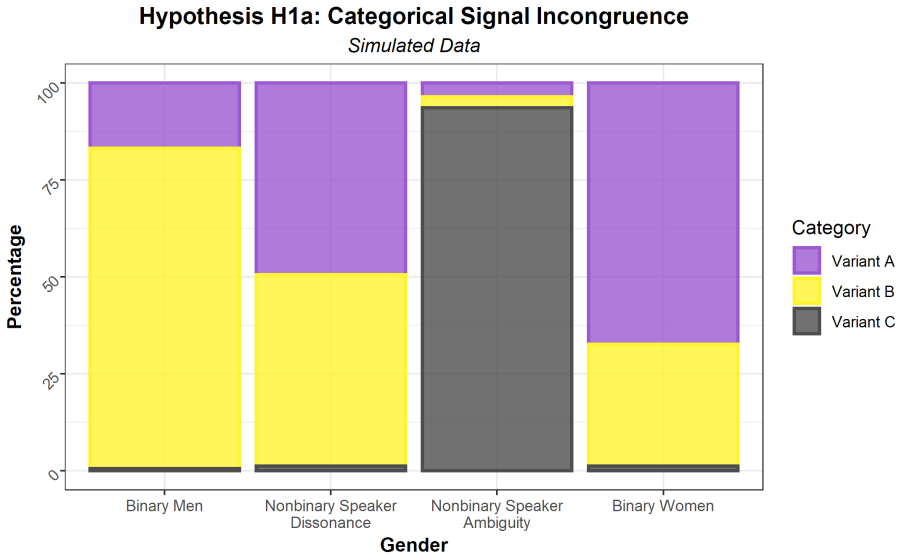


FIGURE 8.1: Simulated distribution of token counts between variants of a given categorical variable. This figure is reproduction of figure 3.1 in Chapter 2. In this case, specificaly regarding intervocalic /t/, the Category can be read as "Production of intervocalic [t]", Variant A (purple) as "fricated tokens", Variant B (yellow) as "flapped tokens", and Variant C (black) as "tokens that are neither fricated nor flapped".

of the dataset and some possible trends in section 8.6.2, I was unable to test this hypothesis due to the paucity of contextual data available.

Hypothesis H3: Influence of the Individual Condition

> *Variation is bound by the individual circumstances of each participant. Individuals with shared socially-encoded circumstances, such as age and binary gender, will speak more like each other.*

H3 predicts that binary-gendered participants will behave more like their same gendered peers, due to similarities in their individual condition. That is, each binary man's production of intervocalic /t/ will be more like the productions of other binary men, and less like the productions of binary women and nonbinary people. Likewise, each binary woman's production of intervocalic /t/ will be more like the productions of other binary women, and less like the productions of binary men and nonbinary people. Further, **H3** predicts that intervocalic /t/ in individual nonbinary speakers will show more variation within-group than that found within binary speakers.

H3 further predicts that individual's use of frication will vary in ways that are consistent with their own identities, values, and histories. Due to the individual condition being, by definition, unique to each participant, it is tautological to make predictions more specific than the group level of broadly shared history: The knowledge required to form a prediction about an individual's behavior would be necessarily based on observing and experiencing that behavior. Section 8.5 presents a number of spotlights, qualitative analyses that demonstrate the effect of the individual condition. Section 8.5.2, in particular, examines **H3** as it relates to the older triad of speakers: **H3**, in line with the apparent time hypothesis, predicts that members of the older triad will speak differently from the younger participants in a number of ways. The binary members of the cohort are expected to produce variants of intervocalic /t/ that align

with their gender and age cohorts, while the nonbinary member, Alex, is expected to produce variation that reflects the unique intersection of age and gender that shapes their identity.

8.4 Analysis

8.4.1 /t/-Tagging

All instances of intervocalic /t/ appearing inside single words produced by participants were extracted from the RAINBO corpus using Fromont and Hay (2012)'s Language, Brain & Behaviour Corpus Analysis Tool (LaBB-CAT)'s search function. Across the 16 speakers, this resulted in 2,208 tokens, which were then hand-coded using a combination of acoustic, auditory, and visual inspection. For the most part, hand-coding followed the same methodology as outlined in Hay and Foulkes (2016, p. 306), with two exceptions: While Hay and Foulkes identified ten possible categories ([t tʰ ʰt t̪ ɾ d d̪ ʔ h] and elision ∅), I did not find either of the two lowered variants [t̪] and [d̪] within my data. Categorisation followed the process outlined in Figure 8.2, as follows: Tokens were first screened for suitability; being correctly aligned, clear speech without errors, hesitations, or speaker overlap, and bordered by vowels on either side. Once determined suitable, the categorisation started with visual inspection of the spectrogram using Praat to determine specific acoustic properties of each token, such as high-frequency aperiodic noise and visible upper formants. I then listened to each token to make a final decision regarding the type of intervocalic /t/ heard. In the rare cases where the spectrogram and my perception of the sound seemed contradictory, I labeled the token as ambiguous and treated it as unsuitable. Example spectra for each type of closure can be found in Appendix G. Elided (N = 221), glottalised (N = 5), and otherwise unsuitable (N = 540) tokens were removed from analysis. This left 1442 tokens for analysis.

Twelve weeks after rating all tokens of intervocalic [t] among the six nonbinary participants, 10 random tokens from each of those speakers were rated again, to check reliability. When using the original categorisations[1], reliability (measured by an exact match in coding between the two time points) was only 64%, with the bulk of mismatches between tokens with similar presence/absence of high-frequency noise. To this end, the six applicable categorisations were collapsed to two: Fricated (N = 553), indicated tokens that contain some evidence of high-frequency aperiodic noise in between the vowels[2], and non-fricated (N = 889), referring to tokens in which there is a clear (acoustic and/or auditory) non-glottal closure in between the vowels but no aperiodic noise[3]. This category simplification increased the reliability across time points to 90%. Collapsing the fricated token category alone led to an increase in reliability to 76%, while collapsing the unfricated category alone led to an increase in reliability to 79%.

8.4.2 Results

A breakdown of the number of tokens by gender and speaker, and the resulting frication rate, can be found in Table 8.1. Overall, with the exception of one speaker (Kelly, a binary woman and a member of the older triad), frication rates are relatively low, in line with the literature showing an increase in flapping over time. When

[1] NA (including alignment error, ∅, and ʔ), [ɾ], [d], [t], [tʰ], [ʰt], and [h].
[2] [ʰt] (N = 41), [tʰ] (N = 153), and [h] (N = 359).
[3] [t] (N = 17), [d] (N = 230), and [ɾ](N = 642).

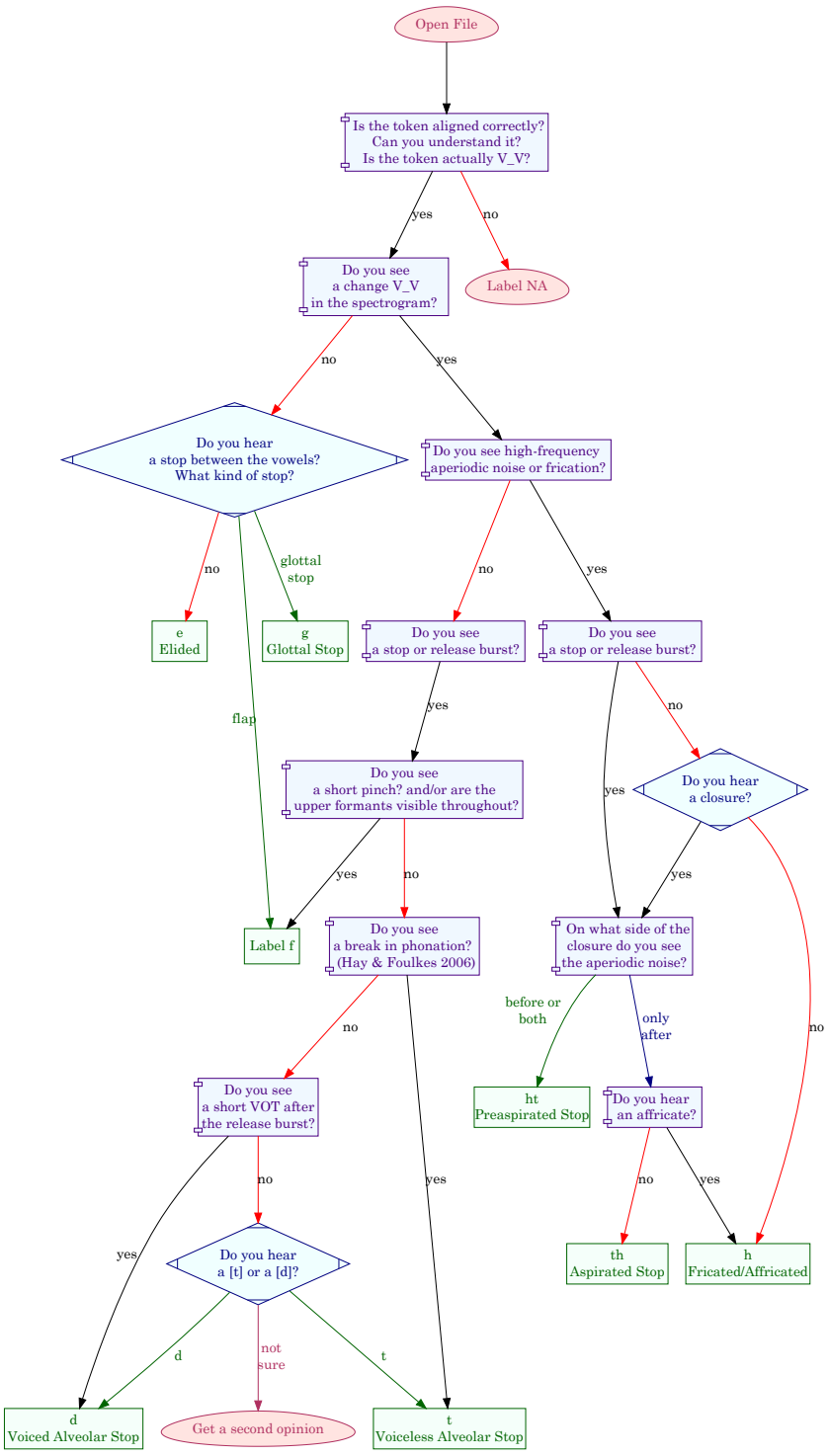


FIGURE 8.2: Flowchart of /t/-determinations. This chart varies from traditional flowchart visual language in that it uses square labels (in purple) to indicate decisions based on visual evidence present in the spectrogram, and diamond labels (in blue) to indicate decisions based on the perceptual information heard when listening to the token.

considering the group frication rate, nonbinary speakers' rates of frication, 38.5%, are exactly between the group averages for binary women (55.52%) and binary men (20%). Individual rates of frication among nonbinary speakers is a mirror image of those found among binary women: Both groups' speakers produce either "low" or "high" frication, with a span of about 30 percentage points between which no speakers fall. For nonbinary speakers, four of six exhibit frication below 30%, and two are above 60%, while for women, one binary speaker has a rate below 20%, and the other four are above 48%.

Gender	Name	Fricated	Unfricated	Totals	% Frication
Nonbinary					
	Mello	8	52	60	13.33
	Ollie	23	129	152	15.13
	Avery	22	79	101	21.78
	Kit	73	174	247	29.55
	Istus	126	61	187	67.38
	Alex	91	53	144	63.19
Group		343	548	891	38.5
Women					
	Reese	15	63	78	19.23
	Quinn	18	19	37	48.65
	Jesse	19	19	38	50
	Remi	35	20	55	63.64
	Kelly	69	4	73	94.52
Group		156	125	281	55.52
Men					
	Devon	6	32	38	15.79
	Taylor	22	104	126	17.46
	Jordon	4	17	31	19.05
	Noel	7	29	36	19.44
	Riley	15	34	49	30.61
Group		54	216	270	20

TABLE 8.1: Intervocalic /t/-frication counts per speaker and gender group.

To determine the extent to which participant gender predicted the production of intervocalic /t/, I ran a generalised linear mixed effects model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2020), utilising the bobyqa optimiser[4] and specifying family as binomial (due to the coding of intervocalic /t/ of each token as a binary value specifying "fricated" or "not fricated"). The model was fit using stepwise backwards regression, in which all interactions of all factors are included in the initial model. At each stage, every simplification involving a single factor is compared to the initial model via a liklihood ratio test, and the factor that contributes least to the overall variance, as determined by the p-value, is dropped. Once any amount of simplification of the model results in a significant loss of explanatory

[4]Use of the default optimiser resulted in the model failing to converge, but further testing via the allfits() command suggess this failure warning was likely a false positive, as all other optimisers, including bobyqa, allowed the model to converge.

power, the stepwise simplification stops and that model is assumed to be the most parsimonious.

The following terms were included in the initial model and subsequently removed in the course of the stepwise model evaluation:

- **REPETITION:** How often a word had been repeated in the previous and current utterance.
  - × *Reason for Removal:* Model failed to converge, due to repetitions being rare and highly correlated (if a word was repeated once, it was likely to be repeated multiple times).
- **STRESSEDSYB:** Whether the /t/ was at the beginning of a stressed syllable.
  - × *Reason for Removal:* Model failed to converge, due to the category containing less than 60 tokens, over half of which were the numeral words *thirteen*, *fourteen*, and *eighteen*, spread evenly across participants.

The final model includes the following fixed effects:

- **TRIGEN:** The participant's gender, coded as nonbinary, male, or female.
- **WORDFREQ:** A scaled measure of word frequency. Before scaling, values ranged from 0 through 357,737. After scaling, values ranged from -0.343 to 28.70.
- **SPRATE:** A scaled speech rate, measured in syllables-per-second at the phrase level. Before scaling, values ranged from 0.962 through 16.667. After scaling, values ranged from -2.65 to 6.05.

The final model also contains the following random effects:

- **1 | SPEAKER/RECORDING:** This random intercept accounts for individual differences across participants and recording number.
- **1 | TEXT:** This random intercept accounts for individual differences for each word.

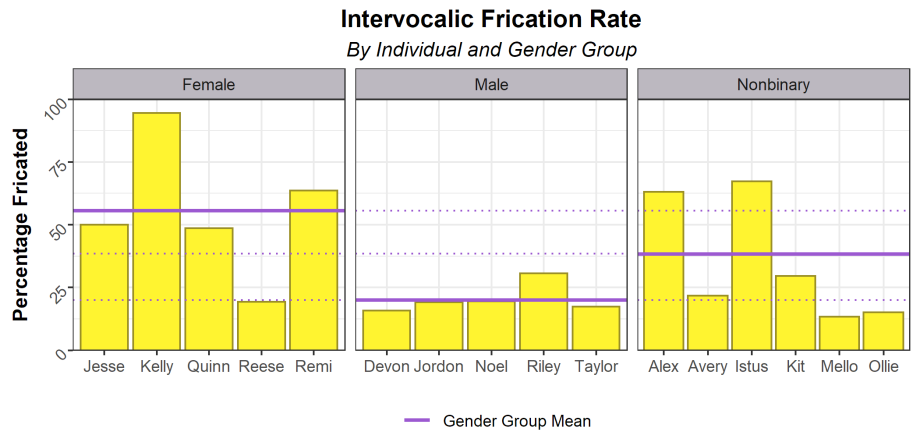
Predictor	Estimate	SE β	z-score	p	*
(Intercept)	−0.648	0.682	−0.951	0.342	
TRIGEN:Female	1.779	1.000	1.779	0.075	
TRIGEN:Male	−1.133	1.022	−1.109	0.268	
WORDFREQ	−1.820	0.514	−3.541	<0.001	*
SPRATE	−0.773	0.138	−5.590	<0.001	*
ANOVA comparison without GENDER				0.043	*

TABLE 8.2: Results of the general linear mixed-effects model, and analysis of variance comparison to the same model with the GENDER term removed.

The results of the model can can be seen in table 8.2. The model contains no significant interactions and two significant main effects, of word frequency (WORDFREQ) and speech rate (SPRATE). Full model output can be found in Section E.3, including a releveled version of the model showing that women's rates of frication are

significantly higher than men's, but do not significantly differ from the nonbinary participants. The significant effect of word frequency and speech rate on frication confirms previous research (Fiasson et al., 2016; Holmes, 1995): Faster speech elicits more flaps, as do more frequent words. In regards to gender, it is found that men flap significantly more often than women, confirming previous literature (Holmes, 1994; Docherty, Hay, and Walker, 2006). It is also found that nonbinary speakers' rates of frication, when taken as a group, fall between he rates of binary men and women, though they vary significantly from either group.

(i)



(ii)

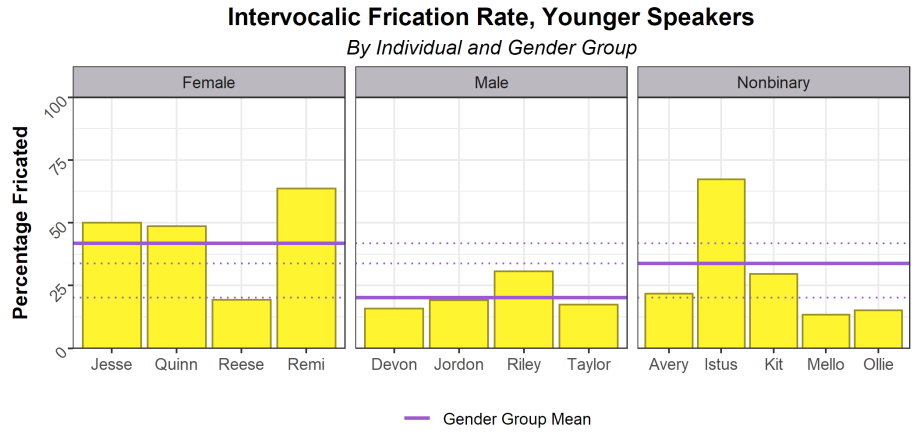


FIGURE 8.3: Percentage of fricated tokens of intervocalic [t], by individual and gender group. All speakers (and group means incorporating all speakers) are included in 8.3i, while 8.3ii have the older triad removed and the means (purple line) updated to reflect that. These updated means show the influence a single speaker can have on the cohort, most clearly in the case of binary women, where Jesse and Quinn both rank above-average frication rates in 8.3ii, once Kelly's data is removed. Note that the total number of tokens measured varies by gender group, and that full token counts and totals can be found in Table 8.1. The older triad's results are discussed specifically in Section 8.5.2.

Figure 8.3 visualises this result, and further emphasises the importance of examining both individual and within-group variation. The variation between individual women is much wider than the variation between groups, or between individuals within the other gender groups. Both Kelly and Alex are producing tokens far above their gender group mean. This is presumably due to their age, since the trend in NZE over time is towards flapping in all cases (Holmes, 1994; Hay and Foulkes, 2016), and Alex and Kelly (along with Noel, a binary male) are members of the older triad. I discuss this triad in more detail in section 8.5.2. After removing these three speakers from the group and looking exclusively at the participants under the age of 25 (as illustrated in Figure 8.3ii), we see that nonbinary speakers hold both the highest (Istus, 67%) *and* lowest (Mello, 12%) overall rates of frication, both of which deserve some comment. I discus Istus' use of frication later in this chapter, as her relationship to particular forms of femininity are likely to be driving this. Mello's incredibly low rates of frication are likely driven by their involvement with a male-dominated sport that involves a lot of travel and international contacts, which presumably includes speakers of other Englishes where flapping is systematic across all genders.

It is also not the case that the nonbinary speakers fall neatly into a bimodal distribution, with some speakers patterning with the binary men and some with the binary women:[5] In fact, when comparing the position of nonbinary speakers in relation to the gender group means, we can see three groups: Speakers fricating below the male averages (Ollie and Mello, while Avery is close), speakers fricating above the female average (Istus and, optionally, Alex), and speakers fricating between binary gender means (Kit and, less centrally, Avery). Of course, analysis is further complicated by the variation within the gender groups: Reese, a binary woman, has a frication rate near (below, in fact) that of binary men, while Riley, a binary man, uses frication at a much higher rate than the others in his gender cohort, though not on par with the rates of the binary women. This result is consistent with findings that men are generally given stricter boundaries when performing masculinity (Pascoe, 2011, p. 66). This also aligns with the stereotypes regarding pitch and pitch range explored in Chapter 2, that show that even within their binary gender cohort, there is evidence that women have access to a wider production space than men.

The results of these analyses show that frication rates among the binary participants in RAINBO are in line with the established trajectory in the literature regarding this variant in NZE. Overall, binary men flap more than binary women. As a cohort, nonbinary speakers differ from binary speakers and from each other: While some nonbinary speakers' rates of frication are similar to those found among binary men or women, sometimes even exceeding the limits found among the binary gender cohorts, others are more centralised. Finally, these results show that intervocalic /t/ frication is not an exclusive marker of binary gender, and is subject to the influence of the individual condition, as evidenced by "atypical" frication rates among binary speakers relative to the other members of their cohort.

[5] I do not provided the Assigned Sex at Birth (ASAB) of speakers here, but it is not the case that ASAB aligns with placement above/below the mean or anything so straightforward. Avery and Istus share an ASAB, for example.

8.5 Spotlights

8.5.1 Introduction

In this section, I highlight a number of ways the use of frication varies among the individual participants in this study. Each spotlight alone represents a small but compelling starting-point for future research. Taken together, they begin to paint a much larger picture about the ways even a single categorical variable contains vast potential for use and (re)interpretation in identity-building.

I first examine the older triad, as they represent a unique subgroup of participants in regards to the differing stereotypes about frication among their generation of speakers. I then highlight Istus' use of frication as a live encoding of her positionality to femininity. Finally, I examine Kit's frication across recordings, as they relate to changes in listener, and how in cases of exact overlap it can be difficult to determine the exact contextual properties that are driving variation.

8.5.2 Spotlight 1: Frication in the Older Triad

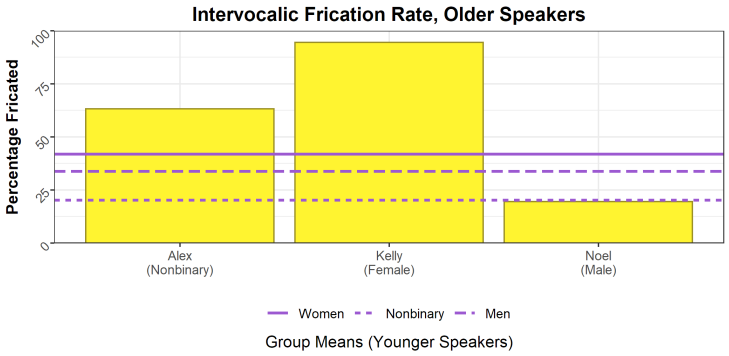


FIGURE 8.4: Percentage of fricated tokens of intervocalic [t], by individual and gender group, older triad only.

Considering the use of flaps in intervocalic /t/, it is useful to turn to the older triad of participants: Alex (47, nonbinary), Kelly (44, female), and Noel (52, male). A direct comparison of their frication of their frication rates can be found in Figure 8.4. Given the flapping of intervocalic /t/ was considered a levelling change in progress at the age of acquisition of these speakers (Holmes, 1994, p. 205), I expect the gender differences to be stronger in the binary members of this older cohort. This is indeed the case, with Kelly's rate of frication being nearly categorical, and much higher than everyone else, including all the younger binary women. Noel's rate of frication, while not the lowest recorded, is still below that of most nonbinary speakers and all of the binary women except Reece (who fricates 0.21% more often). Alex, the nonbinary member of this triad, shows a rate of frication (63%) that is firmly between their two binary age-matched peers, but still higher than almost every other speaker in RAINBO (The two exceptions are Remi (64%), whose frication rate is similar to Alex's, and Istus (67%), whose rate is a little higher, but more on Istus in the next section). That is, when compared to the entire group, Alex's rates of frication are aligned with those found among binary women. But when examined in the context of their age, and compared to other members of the older triad, Alex's use of frication is firmly in line with a strategy of creating incongruence via dissonance.

To help explain this rate of frication, consider this excerpt from Alex's interview:

> "I didn't know there was any other option, so therefore I just went oh well I'm female, but I just redefine what female is to be what I am and it's not the same as what most people label it as [...] It's hard to know how you change yourself, isn't it?"
>
> *– Alex, Interview, 3:20.*

The historical data evident in the literature show that Alex and the other members of their age cohort have a different template for what intervocalic frication *means* in regards to gendered presentation, compared to the other participants. As exemplified in the quote above, until the recent past, Alex's relationship to gender has not been about finding a new, more correct label for their nonbinary identity, but of redefining the label of "woman" imposed on them by the social foundation, and adapting their concept of femininity to include their nonbinary self. Their approach to presenting their identity consequently involves an interpretation of speech that differs from the younger nonbinary participants not only in terms of the shift in stereotypes across generations, but in the sheer time depth of their relationship to those stereotypes.

Alex's description of their past attitude towards femininity make it a real possibility that their past rates of frication were closer to feminine norms (like Kelly's). Alternatively, it may be the case that their frication rates have always held, due to their redefinition of "female" to include their own nonbinary self, whether they had a name for that identity or not. Regardless of whether and when Alex's current way of speaking came to be, their performance of gender is one which takes the binary femme template as the baseline to which their speech is a redefinition, whether the redefinition be "nonbinary" (as it is now), or "a different kind of woman" (as it was in the past).

Precisely because we lack the means to go back in time to examine the old data with a new understanding of gender, the incorporation of nonbinary genders in current and future work is crucial. However, we should not forget that it is also possible, in some ways, to incorporate an expanded understanding of nonbinary gender in regards to *past* research, both through direct study (as I have attempted here and in the previous section on NEAR/SQUARE) and in recognising the potential confounds inherent when the research methodology erases categories by design. Examining these types of changes in language over time among nonbinary people, including those of varying experiences with binary stereotypes, is a worthwhile direction for future research.

8.5.3 Spotlight 2: Istus' Frication in Context: Time

In this section, I briefly highlight two examples of Istus' online changes in intervocalic frication, where she uses flapped variants in contexts where she is positioning herself as liminal to, or partially outside of, femininity. This provides further evidence of Istus' control of her speech identified previously in Chapter 6, where I examined her online changes in pitch for the purposes of embodying (or disembodying) femininity to serve communicative or affective ends.

In this first example, Istus is talking about her relationship to femininity[6]. I have collapsed all categories here according to the criteria outlined in Section 8.4.1. Those labeled *H* indicate tokens of intervocalic /t/ produced with some amount of high-frequency noise, while those labelled *D* indicated tokens containing a non-glottal closure but without noise present.

[6]For comparison, Istus describes masculinity as "Basically the least relatable thing on the planet, after hyenas." (1:25:41).

(3) *Istus Interview Recording 1:27:10-1:27:41*

1 I d~I do the whole nurturing protector thing
H

2 and so like

3 i~back to the like traditional dichotomy of masculinity and femininity
H ***H*** ***H***

4 that I talked about

5 uh it's like yes all the femininity for me please thank you very much
D

6 I'm here to stay . stay

7 but you know

8 but the societal expectations that come with that just like
D

9 no no thank you goodbye

10 it's the femininity without the requirement to perform femininity
H ***H***

When Istus flaps in line 5, she is slightly affecting a voice to represent her inner monologue, though it is not so extreme as the examples in Chapter 6. Still, she positions herself here as someone lacking femininity, in so much as she desires more of it, but also aware of the more negative aspects that come with it, some of which she chooses to reject. This parallels her description in Chapter 6 (6.5.5), where she demonstrates her "natural" voice ("This is what my voice naturally sounds like...") as encoding more masculine features than her "normal" voice (...This is how I normally try and talk.). Her "normal" voice in this excerpt has the feminine-coded fricated intervocalic /t/, while her inner monologue has the more masculine-coded flaps. The content of this excerpt also mirrors her use, and rejection of, the "California Voice", in that Istus does not attempt a wholesale adoption of all feminine stereotypes, but only those aspects of femininity that resonate with her whole sense of self.

In the second excerpt, Istus is talking about her childhood, in particular the messaging surrounding the differences between boys and girls.

(4) *Istus Interview Recording 1:27:10-1:27:41*

1 and you like you go to school and it's like uh boys have cooties
D

2 you know

3 but no cooties don't exist because men and women are
H

4 fundamentally interchangeable

5 ah cooties
D

6 nooo

When Istus affects the "past" voice of talking about cooties from the perspective of her childhood in line 1, she uses a flapped production, but then fricates when she responds from the perspective of her present, adult self. Istus' tone during this shift in temporal perspective imbues the exchange as something of a dialogue with her past self, with lines 3-4 delivered as a gentle, if condescending, presentation of fact, the way you might correct an actual child. In line 5, Istus returns to this "past" voice. The "ah" is an obvious expression of disgust that might be more correctly transcribed as "ugh" or "eaurgh" – exactly the sound you would expect someone to make if they discovered they had touched something inundated with cooties. Line 6 is a long, drawn out groan, where Istus acted out dying from a cootie infection. This minor example adds an interesting perspective to what Hay and Foulkes refer to as speakers' potential to use variation as a way to "portray themselves as historical versions of themselves" (Hay and Foulkes, 2016, p. 323). In the process of (re)creating and (re)iterating one's identity over time, trans and nonbinary people have the potential to provide a much starker contrast between the "past" and "present" self than other populations, as Istus has done here.

8.5.4 Spotlight 3: Kit's Frication in Context: Interlocutors

I examine here another nonbinary speaker's contextual uses of intervocalic frication. Figure 8.5 shows Kit's frication rates across recordings, with the context divided between whether the other speakers in the conversation fall into one of two camps: Whether they are predominately members of Kit's immediate family, or not. While any interpretation should be tempered by the low number of tokens, it is the case that Kit's rates of frication are higher in every recording where their family members are present. Because of the nature of the In the Wild (ItW) recordings, there is difficulty in forming a direct comparison between one participant and the others, as the collection of recordings provided reflect their individual circumstances and differ from each other quite dramatically. While it would be ideal if there were another participant, or two, with a similar division of recordings, it is still illuminating to examine Kit's variation in frication by interlocutor, especially in context with their self-description of their emotional state during each recording.

The division of interlocutors is not completely perfect: ItW recording 5 also includes the presence of strangers in addition to family members. This illustrates some of the complexity involved in determining the sources of this variation across contexts. Kit is out to both their family and friends as nonbinary, so "being out or not" is an unlikely source of this variation (with the possible exception of ItW 5, though Kit says the family guests were likely unaware of their gender as nonbinary). As a stranger, Kit's relationship to me (as a researcher), and the conversational context of "answering questions" is much closer to the context including their parent's guests in ItW 5. It might be that Kit's frication raises when speaking to family for the same reason Istus' frication drops when discussing her childhood: Events (or situations) that evoke a past version of the self may likewise evoke "older" versions of a person's speech, and Kit's experiences with family include their whole life, including the majority of it that predated their identification as nonbinary.

Previous work by Gratton (2016) has established the potential for phonetic variation (in the form of (ING) production) to be affected by a nonbinary speaker's risk

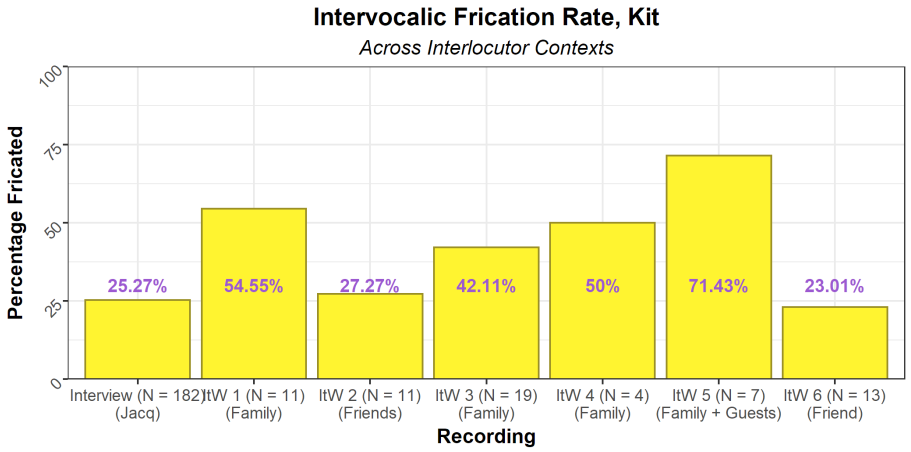


FIGURE 8.5: Kit's frication rates across recording and interlocutor.

of being misgendered, though in that case the direction of the change was towards increased use of the variant that violates the anticipated assumptions of the potential misgendering agent. As this would predict *more* flapping on Kit's part, it is unlikely to be the case here that Kit's variation is rooted in a perceived risk of being misgendered. I propose, instead, that Kit's variation might be an indication of comfort. When Kit describes their mood during the recordings, they use either "tired", "stressed", or both for all recordings apart from ItW 2. During the interview, I asked Kit to describe their emotional state during each recording ("How did you feel during this conversation?"). Here are their responses in reference to the recordings with family:

ItW 1:

> I was tired from being around people all day.
>
> – *Interview 09:33*

ItW 3:

> I was having a very bad day where I was absolutely exhausted and so like, being surrounded by people who I didn't know [...] it was just mum [in the conversation] but there were a lot of people present.
>
> – *Interview 34:40, 35:55*

ItW 4:

> I was very tired, because I think I'd just come back- I'd either come back from work or come back from a social thing.
>
> – *Interview 53:34*

ItW 5:

> Not stressed, but slightly on edge, because mum was bustling around in the kitchen and I had to talk to two people who I didn't really know.
>
> – *Interview 1:06:23*

Compare the above with Kit's description from ItW 2, at a party with many of Kit's friends present:

> It was a bit awkward because I knew I was recording myself [...] I didn't feel uncomfortable but I knew that [the recording equipment] was there.

– Interview 21:30[7]*, 22:24*[8]

and ItW 6, a conversation with a friend at university:

> I was very tired, because I'd been working on the essay, but other than that it was pretty chill.
>
> *– Interview 1:20:01*

The recordings containing family all describe a similar form of social fatigue or anxiety. While Kit also reports being tired in ItW 6, the source is schoolwork rather than social obligation. ItW 2 is the only recording where Kit does not use the words "tired" or "stressed" to describe their mood, and while there were multiple people around, Kit was surrounded by friends and in an environment they identified a few times as comfortable. While I cannot know exactly how stressed Kit was during our interview, I definitely had the impression we were on friendly terms. They came across as quite comfortable, though dynamic, throughout the interview, and even agreed to turn the recording equipment back on a few minutes after wrapping up to talk about their *itabag*.[9] Even given the limitations of the data, Kit's rise in frication seems to coincide with a rise in social fatigue. In the small sample of ItW recordings from Kit, that heightened social fatigue also coincides with interactions with family. With the limited sample of data and contexts, though, it is impossible to know the relationship or directionality, if any, between these factors. If social fatigue elicits frication in Kit, it is unknown whether speaking with family heightens social fatigue, or if Kit was more likely to record with family around when they were already feeling tired. Alternatively, if speaking with family elicits more fricated forms, the presence of these gendered forms, in their speech may contribute to or heighten Kit's social fatigue, whether the pressure to produce fricated forms are rooted in past experiences in this context, explicit (or implicit) gendered expectations on the part of their family, or something else.

8.6 Discussion & Assessment of Hypotheses

8.6.1 H1a - Categorical Signal Incongruence

As a group, it is not the case that nonbinary speakers are statistically more likely to produce flaps than binary women or statistically less likely to produce flaps than binary men, though this is indeed the trend when looking at the summary statistics and macro-level analysis. It is the case, however, that women's use of frication is significantly different than that of men, when they are used as the reference level ($p < 0.01$), confirming previous research on binary populations that link higher frication rates to women (Holmes, 1995; Fiasson et al., 2016). When examining the frication rates of individual speakers, there is evidence that fricated intervocalic /t/

[7] Yeah it was it was a bit awkward because I knew that I was recording myself but we were talking about memes so it was like, it wasn't, like, awkward as in *oh no I don't want to say anything* it's like *oh my god I can't believe I'm saying this on a recording*.

[8] I was trying to like, keep the laughter to a minimum because I was sitting there like, *I know I laugh really loud I don't need to do this on a recording* but it was very, mmm, quite comfortable but like, still, I can't phrase it very well but like I didn't feel uncomfortable but I knew that it was there and because it's still, you know, just the second one.

[9] 痛バッグ – A backpack with a clear window pocket and foam backing, whose purpose is to show your love and dedication to a fandom or character by filling the window with keychains, charms, and other objects. I have more recently seen this type of bag described and sold to Western audiences as a "pin trader" bag.

does not simply encode the sex/gender of femininity, but in fact encodes a complex combination of traits that evokes a *specific* femininity that is both prestigious and mature.

The overall dearth of tokens and the low reliability ratings between tags with similar features necessitated collapsing the instances of intervocalic /t/ into two broad categories (fricated and unfricated). This makes an analysis of gender incongruence via ambiguity impossible, because there is no measurable "third option" available. However, the groundwork is present for future projects examining more tokens, or different categorical variables (for example, variation in ɪn/ɪŋ (Gratton, 2016; Rechsteiner, 2021), pronoun use (Storoshenko, 2021; Loughlin, 2021), concessive clauses (Mondorf, 2002), or hedging (Holmes, 1987).).

The spotlight on the aged-matched triad of speakers over 40 years old patterns very closely to the hypothesised distribution for categorical variables. While three speakers given three gender categories cannot be taken as anything robust, Alex's use of frication as between that of their binary gender peers does suggest the use of incongruence via dissonance, as predicted by **H1a**.

It should also be reiterated that these hypotheses predict variation, not homogeneity, among nonbinary participants, as not all hypotheses can hold true for all speakers at all times. Though the nature of the dataset and the statistical tests involved treat "nonbinary gender" as a third category to allow comparisons to binary men and women (and, by extension, establish the "gendered norms" that it is assumed all nonbinary participants are working within), the coherence of "nonbinary" as a monolithic gender class is superficial. This is evidenced clearly by the range of answers given by participants when asked to self-define their gender, as well as the variation in frication production. Riley's outlying rates compared to other binary men, and the range of frication found among age-matched binary women should not be ignored, either. The latter in particular challenges the coherence of "woman" as a monolithic gender class itself.

8.6.2 H2 - Contextual Incongruence

There is not enough data to address **H2** comprehensively, due in large part to the lack of transcribed interview data for binary participants. If I remove recordings with $\leq$10 instances of intervocalic /t/, every nonbinary participant has at least two recordings to compare across, but only 3/5 binary women have more than one recording left to analyze. The data from the three binary women are as follows: Remi's frication rates range from 46-68% across two recordings; Kelly's frication rate varies from 82-100% across four recordings, and Reese's frication rate varies from 11-26% across three recordings. The binary men have even less data to work with: Ance removing recordings with less than 10 tokens, Taylor is the only male participant with more than one recording left – across his four suitable recordings, the frication rate varies from 11-24% (incidentally, this is almost identical to Reese's frication rate). Of the measurable binary participants, the maximum difference in frication rate, the largest difference between the recording with the lowest frication rate and the recording with the highest frication rate, is 20%, while we have seen in the previous analysis that Istus and Kit can vary by 30% or more across contexts, even when recordings with $\leq$10 tokens are removed. While this, in conjunction with the Spotlight data, provides compelling evidence that nonbinary speakers exhibit considerable contextual variation in their use of /t/ frication, a much more robust examination of a larger dataset, particularly data from binary speakers, is necessary

to determine whether or not context has a greater effect on nonbinary speakers than binary speakers.

8.6.3 H3 - Individual Condition

Binary speakers', in particular binary men's, rates of intervocalic /t/ frication support the predictions made in **H3**, with some caveats. Overall, same-binary-gender groups do fricate at rates close to their same-gender peers. There are two exceptions to this trend, one in each gender group: Riley, a binary man, fricates more than other binary men, in some cases twice as much; Reese, a binary woman, fricates at less than half the rate of the other binary women. While Riley's rates of frication do not reach the expected rate for binary women (discounting Reese), Reese's rates are completely in line with the expected rates for binary men. These exceptions oppose the group-level predictions made regarding **H3**, which are assumed to arise as a result of the greater level of shared experience in the individual condition among same-gender binary participants, especially in regards to the templates for gendered speech. This indicates that, even among "gendered" variables, there is increased room for nuance in investigations for patterns of use not just across gender groups, but within gender groups.

Regarding the nonbinary speakers, the Spotlights have provided evidence for the influence of the individual condition from a qualitative perspective: Alex's use of frication is intertwined with their age, both from the perspective of the older triad's stereotypes differing from the other participants, and from Alex's own historical rapport with femininity. Istus' frication drops when she positions herself as being outside femininity, looking in. For Kit, frication coincides with family, and potentially with social fatigue. For all three nonbinary speakers, their complex relationship with femininity and their own nonbinary gender is reflected in their use of speech markers associated with femininity, like frication.

8.7 Conclusions

8.7.1 Future Directions

The way that I approached the analysis of intervocalic /t/ required the use of a multi-step process with subjective decisions made throughout. To that end, there are multiple possible changes or additions that would result in a dataset from which more reliable evidence can be obtained, and/or more able to handle the nuance apparent in the production of fricated intervocalic /t/. Early on, while I was in the process of collecting and transcribing data for RAINBO I made the decision to prioritise transcribing ItW recordings before other recordings, and from nonbinary speakers[10] over binary men and women. As outlined in Chapter 4, this means that the sociolinguistic interviews from the ten binary men and women are only partially transcribed, and not tagged for intervocalic /t/ at all. Including the binary speakers' interview data would allow a more detailed examination of **H2**, regarding the interaction of gender and conversational context. In the hand-tagging of each token, I sought to maximise the reliability of category membership firstly by prioritising production over perception (via visible features of the spectrogram), and secondly by collapsing similar categories. It would be equally valid to focus on the perception of

[10] I include myself among nonbinary speakers, as the question of accommodation has loomed ever-present in the background, though it is as-yet-unaddressed.

frication, or seek to increase the reliability via means other than category collapse, such as through the use of multiple human raters or a classification algorithm such as random forests (as in Villarreal et al. (2020)'s work on rhotics in NZE). One could also eschew shared categories altogether, and examine the use of frication from a purely speaker-internal baseline: For example, taking Kelly's near-absolute (95%) frication rate as a starting point to examine the way her productions vary across contexts in the duration or location of the high-frequency noise, etc. The scope of the analysis could also be expanded, either to other contexts such as prepausal or across-word, or to other stops such as /k/, where frication might be found to vary.

8.7.2 Conclusion

This chapter explored some of the ways intervocalic /t/ frication is utilised by speakers in presenting a gendered identity. It addressed the potential for nonbinary speakers to create gendered incongruence, particularly in the form of dissonance across signals as predicted by **H1c**, though the evidence for within-category incongruence of either dissonant and ambiguous form is unsupported. It presented a number of small-scale "Spotlights" of the older triad and individual speakers as evidence of the importance of the individual condition on tempering variability. In examining intervocalic /t/ frication across binary and nonbinary speakers, this chapter found that there is more than sex/gender or an encoding of general "femininity" in this and, presumably, other variants. The next chapter builds on this complexity, incorporating the findings from this and the previous analytical chapters to examine the way gender can be constructed acoustically from the interplay of multiple signals at once.

Chapter 9

H1C: Multiple Signals

9.1 Introduction

This section serves to address hypothesis **H1c**, regarding incongruence arising as a consequence of the interaction across multiple signals. I restate the hypotheses and predictions in section 9.2, and discuss the analysis in 9.3. I address the predictions first by examining tripartite and pairwise comparisons in section 9.4, then using the entire data set scaled along a single range in Section 9.5.1 and a dual range in Section 9.5.2. I also directly compare the variation across signals found in two nonbinary speakers, Ollie and Kit, in reference to mean values of binary men and women. Section 9.6 provides two speaker Spotlights, exploring Avery's genderfluidity in context of their relationship to masculinity, and the relationship between Istus' variation in frication with her variation in physical presentation via the use of makeup. I discuss the overall results and how they partially support **H1c**, including some shortcomings and benefits of this type of higher-level analysis, in Section 9.7. I conclude in Section 9.8, summarising how the results from multiple signals tie into the larger account of incongruence presented in the rest of the thesis.

9.2 Hypotheses

Hypothesis H1c – Multiple Signal Dissonance – *Given multiple signals, nonbinary speakers can communicate identity by producing incongruence* ***across*** *signals.*

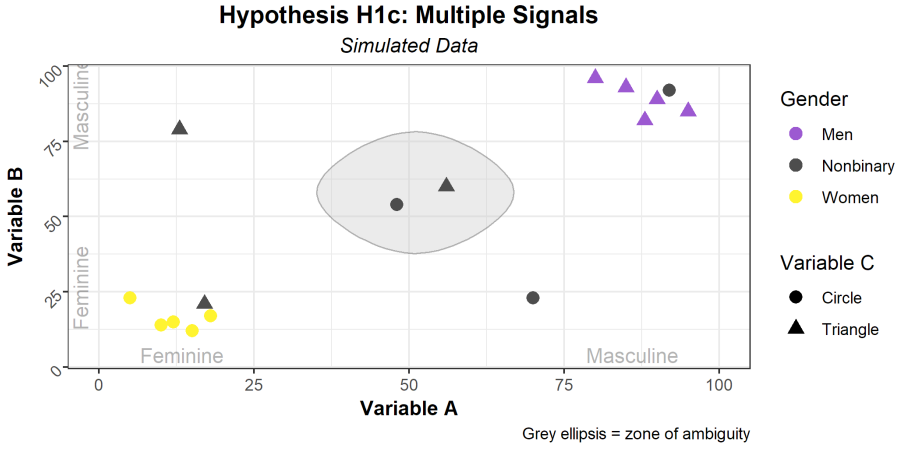


FIGURE 9.1: Simulated distribution of multiple signals (**H1c**), reproduced from Chapter 2.

Figure 9.1, recreated from Chapter 2, illustrates a hypothetical dataset of participants according to three variables, two of which are gradient (A and B, plotted along

the x- and y-axis) and one of which is categorical (C, represented by shapes). These variables are assumed to be exclusively binary, such that more feminine productions are *de facto* less masculine. **H1c** presumes that binary speakers will present overall gender "congruence", in that their productions of variables will generally align with the expected production space of their binary gender, and therefore the examination of multiple signals together will also display this coherence.[1] Nonbinary speakers may create incongruence via dissonance *between* variables by producing some within expected norms for women, while others are produced within expected norms for men. In Figure 9.1, this is represented by speakers who fall within the four quadrants, representing signals that, when viewed in isolation, align with binary-gendered productions (represented in the visualzation as the purple and yellow clusters), but together create incongruence via dissonance. While I focus almost exclusively on the potential for incongruence across signals in the acoustic domain (with the exception of a brief foray into the visual domain in Section 9.6), there is nothing inherently limiting the modality involved in the hypothesis. In fact, multiple participants assert that visual signals are coded so strongly for a single binary gender that they can shift interlocutor perception disproportionately.

> The first thing, people just look at your appearance and they see a man or a woman and if they're a bit more aware they maybe see there's something else but ... I think everyone just makes that kind of split second decision when they see someone.
>
> — *Alex, Interview, 51:23*

> Most cis people tend to take everything at face value.... literally all I need to do is put on a dress and a bra and I get read as female but if I do anything short of that it's like *"Oh, you're a man."*
>
> — *Istus, Interview 1:13:12*

9.3 Analysis

I address **H1c** first via pairwise comparisons of each variable, and second by examining all variables together. The pairwise comparisons involve a relatively straightforward, non-statistical analysis whereby all participant's productions are visualised with a single continuous variable on the x-axis and a single continuous variable on the y-axis. This differs from the theorised comparison in Figure 9.1 in that only two variables are plotted at a time, rather than a third, categorical variable plotted using shape differences. While I did analyze a categorical variable, intervocalic /t/ frication, the actual use of this variant by most participants is variable, and so I plot this variable as a percentage of total tokens fricated, therefore treating it as continuous.

Instead of reiterating the methodology involved in obtaining each measure, I instead list the measurements used for each variable, the gendered association/direction, and a referral to the chapter in which the more detailed methodology appears.

1. **Monophthongs** (Chapter 7)
    - ▷ This is represented by taking the mean F2 of the four vowels that were found to differ significantly by gender; KIT, LOT, STRUT, THOUGHT. This

[1] This is also why multiple signal incongruence via ambiguity is not addressed in **H1c**, as it is assumed to be an epiphenomenon, arising from multiple instances of within-signal ambiguity. This is represented by speakers whose productions fall in the central area, represented in Figure 9.1 by a grey ellipses.

is quite an abstract measure, but it is important to remember that the direction of variance was the same in all cases, regardless of statistical significance, with the F2 of nonbinary participants overall falling below that of binary women and above that of binary men.

- Higher F2 is considered more stereotypically feminine, lower F2 is considered more stereotypically masculine.

2. **Vowel Space Area (VSA)** (Chapter 7)
   - ▷ This is measured as the total area, in Hz^2, of a convex hull whose points represent the means of the outermost vowels in each speaker's vowel space. This measure was found to show the greatest difference across binary gender.
   - Larger VSA is considered more stereotypically feminine, smaller VSA is considered more stereotypically masculine.
3. **Pitch** (Chapter 6, Table 6.4)
   - ▷ This is measured as the mean value, in Hz, obtained for each speaker from Talkin (2015)'s **R**obust **E**poch **A**nd **P**itch **E**stimato**R**. (REAPER) program.
   - Higher pitch is considered more stereotypically feminine, lower pitch is considered more stereotypically masculine.
4. **Pitch Range** (Chapter 6, Table 6.4)
   - ▷ This is measured as the distance between the minimum and maximum pitch, in Hz, obtained from REAPER.
   - Larger pitch ranges are considered more stereotypically feminine, smaller pitch ranges are considered more stereotypically masculine. Though note this analysis was complicated by the presence of multiple modes among binary female and nonbinary participants, a nuance which is not accounted for in this analysis of multiple signals.
5. **Intervocalic /t/ frication (VtV)** (Chapter 8, Table 8.1)
   - ▷ Measured as the percentage of total intervocalic /t/ tokens produced with frication.
   - For this scale, more frication is considered more stereotypically feminine, and less friction (more flapping) is considered more stereotypically masculine. However, note that, as examined in Chapter 8, frication appears to index as particular subset of femininity, rather than a broad-spectrum indication of binary gender/sex.

Table 9.1 shows the scores used for each variable for each participant in the pairwise comparisons, as well as the mean score of each variable for the binary gender groups. These values are used in both Min-Max scaling procedures, and in the pairwise visualisations. In addition to examining participants' productions of multiple signals in pairwise and tripartite comparisons using raw values, I also applied a scaling measure to normalise across variables in two ways, a single scale and a dual scale, both of which used a Min-Max scale formula $X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}}$, multiplied by 100 to achieve a final scale between 0 and 10. In the single scale each variable is given a single "sex/gender" score, while in the and dual-scale each variable is given a separate

Gender	Name	Variable				
		Vowels	VSA	Pitch	P. Range	VtV
Women		*F2 (Hz)*	*Hz²*	*Mean Hz*	*Hz*	*% Fricated*
	Jesse	1483.730	253 751.2	194.30	177.72	50
	Kelly	1448.097	248 054.6	164.64	241.06	94.52
	Quinn	1378.450	231 695.9	169.28	238.76	48.65
	Reese	1379.721	126 456.5	173.20	256.10	19.23
	Remi	1363.175	322 198.2	178.93	256.10	63.64
Group Mean		1410.634	236 431.3	176.070	233.948	55.208
Men						
	Devon	1303.609	105 324.6	108.68	139.68	15.79
	Jordon	1237.272	152 588.0	114.47	119.85	19.05
	Noel	1157.547	198 084.7	100.38	94.90	19.44
	Riley	1296.803	114 226.0	110.68	152.76	30.61
	Taylor	1283.520	199 817.7	137.74	144.08	17.46
Group Mean		1255.750	154 008.2	114.390	130.254	20.470
Nonbinary						
	Alex	1366.348	178 234.3	184.47	250.59	63.19
	Avery	1205.926	159 902.1	106.01	133.56	21.78
	Istus	1210.912	214 467.9	130.21	154.13	67.38
	Kit	1325.538	309 732.3	195.88	230.75	29.55
	Mello	na	na	193.88	226.80	13.33
	Ollie	1211.614	167 475.0	148.62	218.39	15.13
Group Mean		1264.068	205 962.3	159.85	202.37	35.06

TABLE 9.1: All measures, for all variables, for all participants, including binary gender means. These values are the source data for the pairwise visualisations and scaling throughout this chapter.

score for male and female norms. A single scale measure most closely represents the conditions required for the predictions of **H1c**: Binary gender is conceived of as exclusive and all-encompassing, and any change to a signal along one gender binary results in a proportionate change across the other. That is, a signal that is more masculine than another is proportionately less feminine, and a signal that is feminine is also inherently emasculating. A dual scale measure, on the other hand, attempts to incorporate standards for each binary gender independently, while still recognising their integrated and directional relationship. This deviation from the original hypothesised structure is motivated by both qualitative and quantitative data collected over the course of the thesis. Qualitative data in the form of participants' identification of the disproportionate effect of certain signals, and the nonbinary participants' capacity to identify when this can be manipulated, is supported quantitatively by the disproportionate use of intervocalic /t/ frication to encode particular forms of femininity found in Chapter 8. The use of a dual scale is a rudimentary way to address this disproportionate gendering, in the absence of a more robust measure of proportional signal weighting, such as one supported by experimental data in gendered perception. As scaling is a process of normalisation, the critiques outlined in Chapter 2.2.2 apply.

For the single-scale measure, for each variable listed in Table 9.1, every participant's score was scaled between 0 and 10, with 0 being the most masculine presentation of the variable, and 10 being the most feminine presentation. The association between which "direction" on the scale each binary gender took was determined

by the clear trends in the literature for every variable except F2[2], which used the directionality identified in the present data, of women's F2 being reliably higher than men's in all the vowels of interest. Scaled values can be seen in Table 9.2. Using the intervocalic /t/ frication rates as an example, the scaling process is as follows: Low frication is determined to be more masculine, and high frication is more feminine. As Mello's frication rate is the lowest among all participants at 13.33%, I assigned them a Min score of 0. As Kelly's frication rate is the highest among all participants at 94.52%, I assign her a Max score of 1. Every other participant's frication is scored along this Min-Max scale according to the formula $X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}}$, with the results multiplied by 100 to achieve a final score between 0 and 10. These scores can be seen in the VtV column of Table 9.2.

Group	Name	Variable					Total	Mean	SD
Women		Vowels	VSA	Pitch	P. Range	VtV			
	Jesse	8.69	6.84	9.83	5.14	4.52	35.03	7.01	2.27
	Kelly	7.05	6.58	6.73	9.07	10	39.43	7.89	1.55
	Quinn	5.92	5.83	7.21	8.92	4.35	32.23	6.45	1.72
	Reese	5.7	0.97	7.63	10	0.73	25.02	5.00	4.09
	Remi	7.08	10	8.23	10	6.2	41.50	8.30	1.71
Group		6.89	6.04	7.93	8.63	5.16	34.64	6.93	2.27
Men		Vowels	VSA	Pitch	P. Range	VtV			
	Devon	2.95	0	0.87	2.78	0.30	6.9	1.38	1.39
	Jordon	2.85	2.18	1.48	1.55	0.70	8.76	1.75	0.81
	Noel	1.89	4.23	0	0	0.75	6.92	1.37	1.79
	Riley	2.86	0.41	1.08	3.6	2.13	10.01	2.02	1.29
	Taylor	3.72	4.36	3.91	3.05	0.51	15.54	3.11	1.53
Group		2.85	2.24	1.47	2.12	0.89	9.63	1.93	1.36
Nonbinary		Vowels	VSA	Pitch	P. Range	VtV			
	Alex	5.86	3.36	8.81	9.66	6.14	33.83	6.77	2.52
	Avery	2.81	2.52	0.59	2.4	1.04	9.36	1.87	0.99
	Istus	1.74	5.03	3.12	3.67	6.66	20.23	4.04	1.88
	Kit	6.6	9.43	10	8.43	2	36.45	7.29	3.23
	Mello	na*	na*	9.79	8.18	0	17.97*	5.99*	5.25*
	Ollie	2.14	2.87	5.05	7.66	0.22	17.94	3.59	2.86
Group		3.83	4.62	6.23	6.67	2.68	23.56	4.71	2.30

TABLE 9.2: Table comparing all participants productions of all variables to binary gender norms - Single scale. Totals closer to 50 are "more feminine" and numbers closer to 0 are "more masculine". Note that Mello is missing vowel data, and they are not included in the group totals for columns marked with a *. VSA refers to Vowel Space Area; VtV refers to intervocalic /t/ frication.

To apply a dual scale, the means of each binary gender group were taken as a baseline measure. Ranking each participant's productions in reference to masculine and feminine norms, the mean value for each variable for each binary gender (listed in 9.1) was assigned a Min score of 0. This average production was taken to be a minimum cut-off measuring something like "sufficiently gendered production", such that "more strongly gendered" productions (falling above/below the mean in the expected direction) were also assigned a score of 0. Remaining productions for each

[2]While previous work, as I explored in Section 2.2.1, shows that differences in F2 are a reliable marker of gender, the relative direction can change depending on the language(s) and vowel(s) of interest.

variable were scaled between 0 and 10, according to the Min-Max formula described above, based on their proximity to the mean score. A sixth measure was added to the dual scale, which I call Vowel Space Conformity (VSC), intended to numericise the visual analysis of vowel dispersion/shape in Chapter 7. While a formalized measure of VSC has been used before in acoustic analysis of single vowels or vowel overlap (Haynes and Taylor, 2014; Kelley and Tucker, 2020), I am unaware of it being applied to the entire vowel space as a way to assess cross-speaker differences. To measure VSC, a convex hull consisting of the mean points for all vowels for each binary gender group was obtained. Intersecting geometries between those mean hulls and the speaker-specific hulls were identified using the Simple Features package in R (Pebesma, 2018). As will become apparent, this approach of quantifying Vowel Space Conformity (VSC) resulted in some extreme values, particularly in cases where speaker hulls were created from disparate vowels. For this size of data, an analysis of VSC works best from a qualitative perspective, comparing individual speakers to gender norms directly via visual analysis (as I did in Chapter 7). This approach will also likely work more consistently if applied to larger datasets, or datasets where vowel hulls use a consistent set of vowels rather than using the entire vowel inventory to construct a convex hull. While a relatively poor measure of actual conformity in this case, I retain it here with a note of caution because it aligns with the analysis in Chapter 7, and does highlight some interesting variation among speakers.

Dual scale values, totals, and standard deviations for each speaker can be found in Table 9.3. I explain the dual scaling procedure here, using the same example variable of intervocalic /t/ and the same assumption that more frication is relatively more feminine, and that less frication (more flapping) is more masculine: The mean frication score for men is 20%, and the mean score for women is 55.52%. Jesse's frication rate is 50%, which means that she is assigned a score of 1.31 in relation to the women's mean, and 4.03 in relation to the men's mean. Kelly, on the other hand, with the highest overall frication rate of 94.52%, is assigned a score of 0 in relation to the women's mean (as her rate of frication is higher than average), and 10 in relation to the masculine mean.

Scaling each participant in reference to all other participants, as in the single scale approach, makes individual outliers clear, while at the same time potentially clouding individual differences in non-outlying data. In cases without outliers, conversely, the single scale risks artificially inflating differences where none may exist. The single scale is gender-agnostic on an individual participant level, as can be seen by a nonbinary speaker being assigned the "most masculine" frication score: This is beneficial in that the nonbinary speaker's score does indeed represent the "least fricated" production, but also highlights the misnomer involved in equating "least/most fricated" with "least/most masculine". The dual scale deliberately inflates the difference on one side of each binary mean, while collapsing the data on the other half to 0. The influence of binary outliers in the dual scale, therefore, is tempered in individual scores but shifted at the group level due to using the mean value for each gender group as the scaling point.

These two scales provide two different perspectives on gendered production: The single scale works from the presumption that the shared social foundation treats the property of "sex/gender" as two endpoints of a single spectrum, polarised as masculine/male/man at one end, and feminine/female/woman at the other. In this conception, any reduction in masculinity results in a proportionate increase in femininity. The dual-scale approach, conversely, treats each binary gender as a separate entity, and explores variation from the perspective of orientation towards, or deviance away from, binary norms. The influence of the shared social foundation on

Name	Gender	B. Norms	Variable						Total	SD
			Vowels	VSA	VSC	Pitch	P. Range	VtV		
Jesse	Woman									
		Femme	0.11	0	2.98	0	4.04	1.31	8.44	1.74
		Masc	8.13	6.07	7.59	9.81	3.77	4.03	39.39	2.39
Kelly	Woman									
		Femme	1.12	0	2.75	1.51	0	0	5.37	1.12
		Masc	5.83	5.75	7.46	6.17	8.8	10	44	1.76
Quinn	Woman									
		Femme	1.42	0	1.93	0.9	0	1.63	5.87	0.83
		Masc	4.24	4.81	6.03	6.74	8.62	3.84	34.29	1.79
Reese	Woman									
		Femme	1.79	8.12	0.22	0.38	0	8.6	19.18	4.08
		Masc	4.09	0	2.10	7.22	10	0	23.41	4.05
Remi	Woman									
		Femme	1.37	0	7.31	0	0	0	8.69	2.93
		Masc	5.42	10	10	7.92	10	5.86	49.2	2.15
			Vowels	VSA	VSC	Pitch	P. Range	VtV		
Devon	Man									
		Femme	5.30	10	1.10	8.9	6.78	9.42	41.5	3.35
		Masc	0.88	0	0.25	0	0.75	0	1.88	0.40
Jordon	Man									
		Femme	5.36	5.97	3.00	8.14	8.21	8.64	39.32	2.19
		Masc	0.87	0.27	1.54	0.01	0	0	2.69	0.63
Noel	Man									
		Femme	7.08	2.09	10	10	10	8.55	47.72	3.10
		Masc	0.67	2.88	3.92	0	0	0	7.48	1.72
Riley	Man									
		Femme	5.42	9.24	4.89	8.64	5.84	5.9	39.94	1.81
		Masc	0.7	0	1.33	0	1.79	1.42	5.24	0.76
Taylor	Man									
		Femme	4.40	1.94	3.95	5.06	6.46	9.02	30.84	2.41
		Masc	1.22	2.98	3.45	2.87	1.1	0	11.62	1.36
			Vowels	VSA	VSC	Pitch	P. Range	VtV		
Alex	Nonbinary									
		Femme	1.79	3.39	2.54	0	0	0	7.72	1.5
		Masc	4.15	2.01	5.16	8.6	9.56	5.8	35.28	2.81
Avery	Nonbinary									
		Femme	5.60	5.34	6.63	9.26	7.22	8	42.01	1.48
		Masc	0.77	0.69	1.59	0	0.26	0.24	3.55	0.57
Istus	Nonbinary									
		Femme	7.53	0.69	8.43	6.06	5.74	0	28.45	3.55
		Masc	0.68	3.82	3.27	1.94	1.9	6.36	17.97	1.99
Kit	Nonbinary									
		Femme	1.48	0	6.28	0	0.23	6.16	14.15	3.04
		Masc	4.62	9.29	9.49	10	7.99	1.28	42.67	3.46
Mello	Nonbinary									
		Femme	na	na	na	0	0.51	10	10.51*	5.63*
		Masc	na	na	na	9.75	7.67	0	17.42*	5.14*
Ollie	Nonbinary									
		Femme	6.31	4.7	3.16	3.63	1.12	9.57	28.49	2.92
		Masc	0.09	1.13	2.83	4.2	7	0	15.52	2.73
			Vowels	VSA	VSC	Pitch	P. Range	VtV		
Men's Means										
		Femme	5.51	6.37	4.13	8.15	7.46	8.42	40.04	1.66
		Masc	0	0	0	0	0	0	0	0
Women's Means										
		Femme	0	0	0	0	0	0	0	0
		Masc	5.54	4.23	6	7.57	8.24	4.77	36.40	1.56

TABLE 9.3: Table comparing all participants' productions of all variables to binary gender norms – Dual scale. This scaling takes the mean of the binary participants' values for each measure, and scales "distance from mean" for a score: Lower numbers represent productions closer to binary norms. VSA refers to Vowel Space Area; VSC refers to Vowel Space Conformity; VtV refers to intervocalic /t/ frication.

the relationship between masculinity and femininity is encoded bidirectionally, in how we interpret productions as "closer" or "further" away from baseline gendered norms. These scaling measures were approached from the point of view of allowing a way to compare quite disparate variables in a relatively straightforward way, acknowledging the shortcomings of this kind of normalization process, rather than with the intention to rank the gendered productions of speakers. I do not propose this method to be particularly methodologically rigorous nor would I suggest that it is appropriate for use outside of this limited context, with these particular variables and speakers, and the shortcomings of the process centered as I have done. This scaling allows a point of comparison that would otherwise not be possible, and I believe that it can provide a useful angle of analysis, so long as the limitations in the methodology and the interpretive nature of the following analyses are kept in mind.

9.4 Results: Tripartite and Pairwise Comparisons

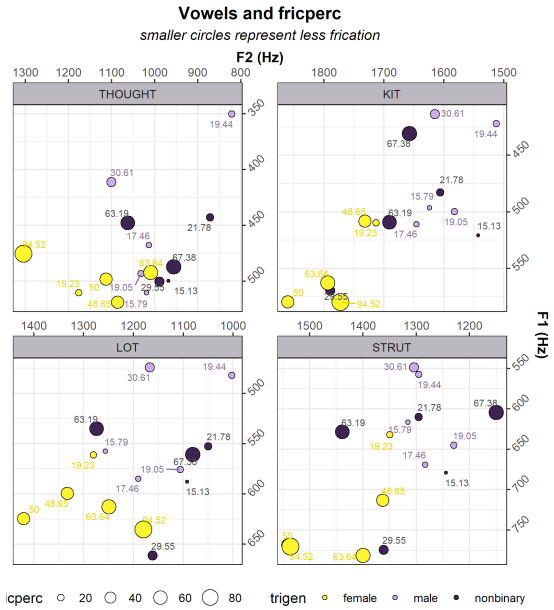
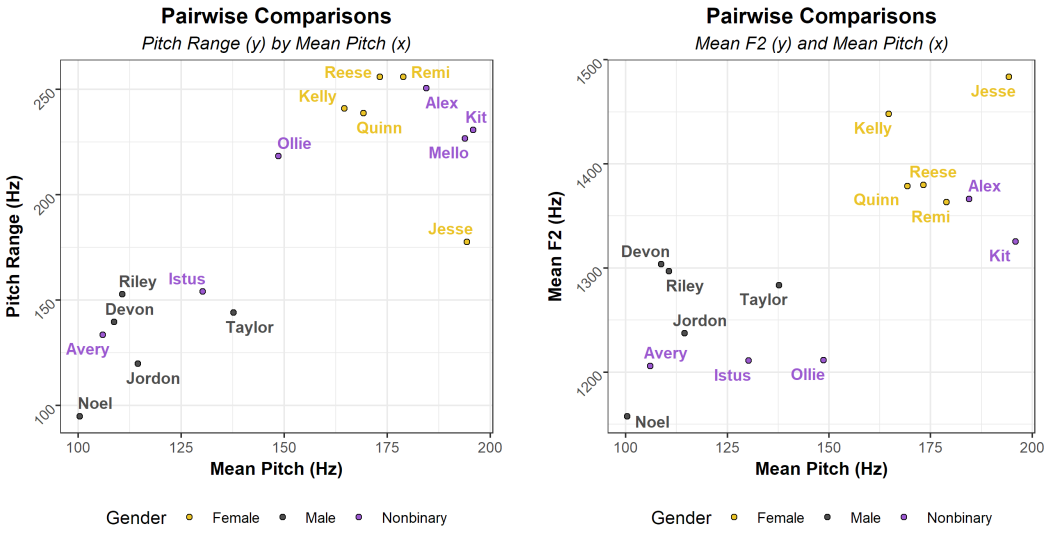


FIGURE 9.2: Vowel production (F1 and F2) by frication and gender group. The size of the bubble represents the frication: larger bubbles indicate greater frication. Frication rates, listed as a percentage of total total tokens fricated, are also indicated beside each bubble. F1 was not found to be significant in the analysis of vowels and is not considered further, and is included as the readability of this visualisation is low, and is increased by including F1 so the general position of the vowel can be discerned.

I present only one tripartite comparison here because, as should be apparent, these visualisations are not very informative when addressing three continuous measures. Figure 9.2 plots the four vowels found to differ significantly by gender in F2. This tripartite visualisation does not tell us anything that is not readily apparent either from the pairwise comparisons below, or aggregate scaled scores that follow, while also containing quite a lot of confusing visual noise. I present this example as an analog to the theoretical/hypothesised visualisation presented at the beginning of the chapter.



(i) Speaker pitch by Speaker pitch range

(ii) Speaker pitch by F2.

FIGURE 9.3: Pairwise comparisons of mean pitch and pitch range (left) and mean pitch and Vowel F2 of KIT, LOT, STRUT, and THOUGHT (right). Note that Mello is absent from 9.3ii as they did not have enough data to be included in the vowel analysis.

Of the possible pairwise comparisons, I highlight only half (5/10) here, that best illustrate differences (or lack of difference) in productions by gender, and the presentation of incongruence. Something these pairwise comparisons make clear is the different degree of "gendering" apparent in speaker productions of each variable. Full pairwise and tripartite results can be found in Appendix H.

Figure 9.3 shows pairwise comparisons representing variables that are strongly gendered along a binary: Pitch range, mean pitch, and vowel F2. Figure 9.3i compares mean pitch (F0) and pitch range, while 9.3ii compares mean pitch with F2. In these visualisations, the binary men and women form two very distinct clusters. The nonbinary speakers also fall into two clusters, though with a much wider range. We can easily imagine a negative slope bisecting the graphs, indicating productions that are equally gendered. This slope separates every speaker, including nonbinary speakers, into two groups, though nonbinary speakers are much closer to the central area, with Ollie crossing the boundary, as his pitch range and F2 pattern differently (while his mean pitch is central). While there is no case where *any* speaker reliably/-clearly falls into quadrants that may represent incongruence via dissonance, Ollie's "boundary crossing" comes close. In subfigure 9.3i, both Ollie and Istus are nearing the central area, as is Taylor, a binary man. The nonbinary participants' productions are consistent with the physiological constraints on production and both Istus and Ollie's self-identification – Istus as "high femme" and also "genderless potato person" and Ollie as "agender". Taylor's productions, while nearing centrality, are still firmly within the cluster of other binary men, and Taylor himself is firmer in his masculine identity than the visualisation might suggest.[3]

[3] "Most male guys probably like sound like me and stuff but like, sometimes other males who don't like to identify as much as males, they typically have a different kind of voice [...] [They] give the assumption that they might not like "*male*", like *being male*. (J: I see, yeah, so they're kind of trying to separate themselves?) Yeah like they're trying to be different to the stereotypical male, they're trying to be their own kind of, either male or like, transgender or something." Interview 53:00

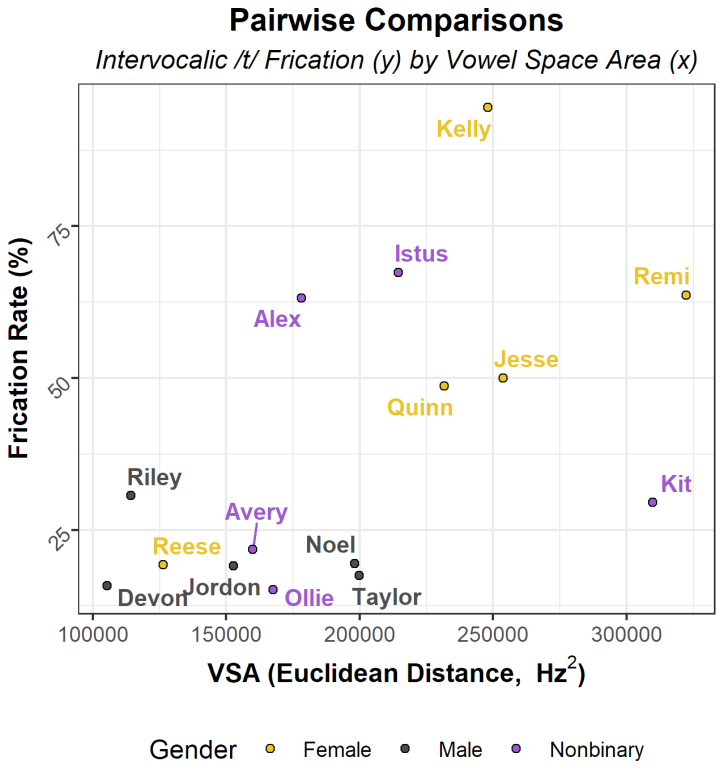
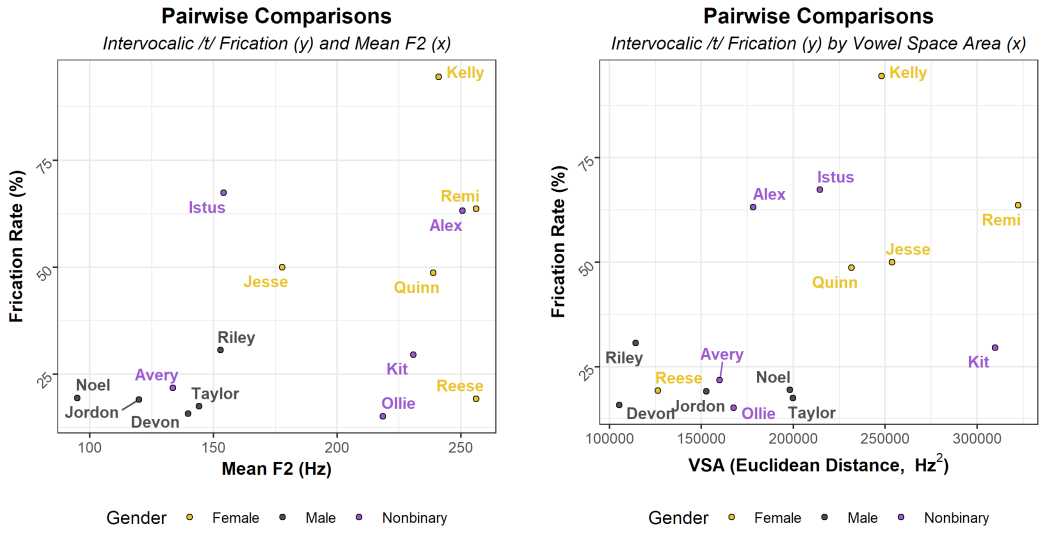


FIGURE 9.4: Pairwise comparison of speaker pitch and VSA.

Figure 9.4 plots mean pitch alongside another, less reliably gendered variable, VSA. The negative slope separating binary gender is less apparent here, as the women's vowel spaces are much more widely distributed. The nonbinary speakers, on the other hand, are spread throughout both axes, with Avery and Kit falling firmly into the same areas occupied by binary peers, while some, such as Alex, are much closer to dissonant production spaces. Reese's vowel space separates her clearly from the other women, while her pitch separates her from the men, Ollie, Taylor, and Istus are once again approaching the central area where we predict incongruence via ambiguity.

Figure 9.5 shows the pairwise comparisons of frication with vowel F2 (subfigure 9.5i) and VSA (subfigure 9.5ii). Frication, as I have shown in Chapter 8, is gendered in a way unlike pitch or pitch range, in that it encodes a particular kind of femininity that does not reflect all female speakers. These visualisations show quite clearly that binary women are able to utilise frication optionally, while binary men do not fricate much at all. I want to draw particular attention here to Istus in 9.5i and Kit in 9.5ii, whose productions here suggest incongruence via dissonance. The unique positioning of intervocalic /t/ frication as not just gendered, but evoking a particular model of femininity, may provide some insight into this variation in nonbinary speakers as regards **H3**: Both Istus and Kit's use of frication aligns with their conception of their nonbinary selves in relation to femininity, Istus in invoking it and Kit in rejecting it.



(i) Speaker frication by F2.

(ii) Speaker frication by speaker VSA

FIGURE 9.5: Pairwise comparisons of frication and vowel F2 of KIT, LOT, STRUT, and THOUGHT (left) and frication and VSA (right).

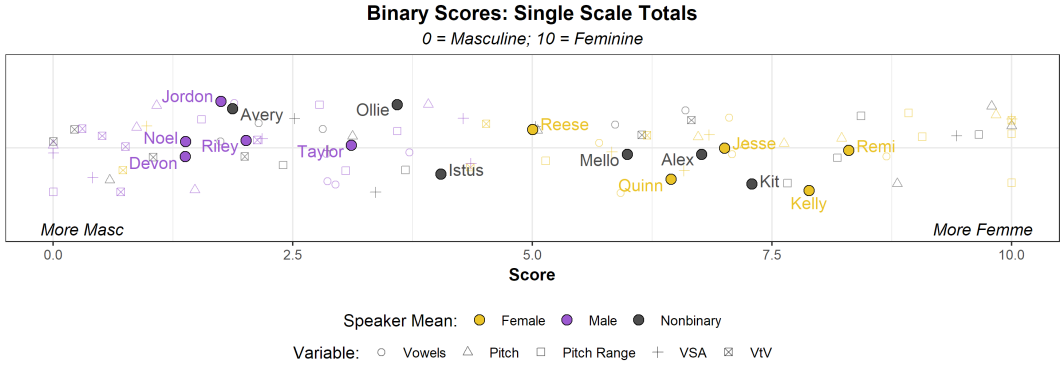


FIGURE 9.6: Single scale scores for multiple signals, plotted for all speakers. Note the y-axis has no scale, and points are jittered vertically for readability only.

9.5 Results: "Total" Incongruency

9.5.1 Single Scale

As pairwise plots for each variable using the singular scaling are identical (except for axis labels) to those already presented and found in Appendix H for the raw values, I do not reproduce or discuss them here.

According to the predictions of **H1c**, when reducing each participant's production of multiple variables to a mean along a single scale, both incongruence via dissonance and incongruence via ambiguity would result in a middling score. Figure 9.6 plots the values listed in Table 9.2 for every speaker on a single plot, while Figure 9.7 presents this same information but faceted by speaker. When examined altogether, as in Figure 9.6, it is obvious that the binary speakers fall into two groups aligned with their sex/gender. These also confirm that the nature of the binary groups is slightly different, with men generally constrained (closer together) and women more widely distributed, while no speaker occupies the "maximally gendered" areas at either

pole.[4]. Nonbinary speakers are spread across and between both groups. The absolute central area (indicating a potential incongruence across signals) is occupied by Reese, a binary woman, along with Istus and Mello. However, shifting focus from the direct center and towards the natural gap between the two groups of binary speakers shows Istus (and Ollie) occupying the central space with Reese more closely representing the "masculine end of femininity", which is more in line with the patterns found in the pairwise comparisons.

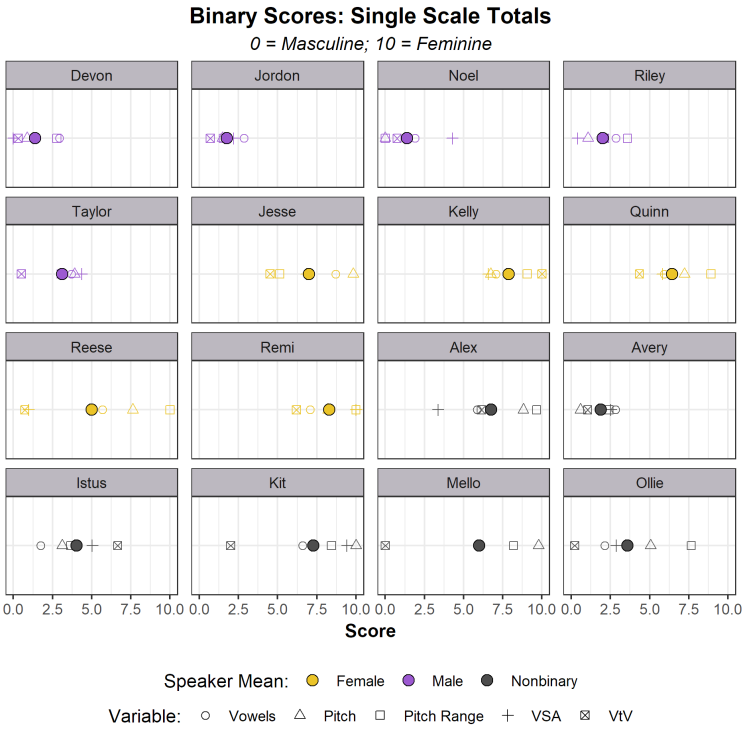


FIGURE 9.7: All variables for all speakers, plotted along single scale. Values closer to 0 are stereotypically more masculine while values closer to 10 are stereotypically more feminine. Note that Mello is missing values for Vowels and VSA.

To better explore how incongruence arises across signals within speakers, Figure 9.7 plots each speaker's scaled mean along with their scaled scores for individual variables. The binary men (and Avery, who I discuss in the following Spotlight), unsurprisingly display the least amount of variation across variables, their productions clustered quite close together. The binary women, with the exception of Reese, tend to cluster in the other half of the score range, though their overall distribution is much wider than the men, which is also represented by the overall larger SDs for most women in Table 9.2. Women having access to a greater range of production than men is the same pattern we tend to find within each variable examined on its own, so it is expected when looking at multiple variables together. The nonbinary speakers show a wider range across variables than either binary gender, though again with the two

[4]While far beyond the scope of this thesis, it is interesting to consider why this might be the case. It may be an indication that speech stereotypes themselves are an exaggerated reflection of reality, and illustrate the limits of attributing continuous scales to proportionately gendered perceptions. It is also possible that there is a perceptual cutoff whereby someone producing speech that is *too* strongly evocative of binary gender may start sounding "unnatural" and so even binary speakers compensate when speaking to ensure they are producing the level of binary masculinity/femininity they perceive themselves as occupying.

exceptions of nonbinary Avery, whose range is very narrow, and binary female Reese, whose range is very wide.

The nonbinary speakers as a group show much less tendency to pattern with each other than the binary gender groups. Istus' productions are generally more centralised than all other speakers, providing further evidence that she may be creating incongruence via ambiguity. Ollie is also somewhat more centralised relative to the other speakers, but not to the same extent as Istus, as his productions trend towards the masculine. (I have previously explored Ollie's relationship with masculinity and gender in Chapter 6.2). Mello's productions align with the hypothesised incongruence via dissonance, but this result must be tempered by their lack of vowel data, and the resulting outsized effect that their low rate of frication has on their mean score. Alex's productions also appear to be somewhat tri/bimodal, with their pitch variables (mean pitch and pitch range) patterning together, while the other variables are dropped. Alex's distribution, however, relative to the other speakers, is between the upper and lower range of the feminine half of the scale rather than across the midway point. Ultimately, it is clear that, while binary speakers tend to show gender consistency across multiple signals in production, nonbinary speakers do not, at least not to the degree that the binary speakers do.

9.5.2 Dual Scale

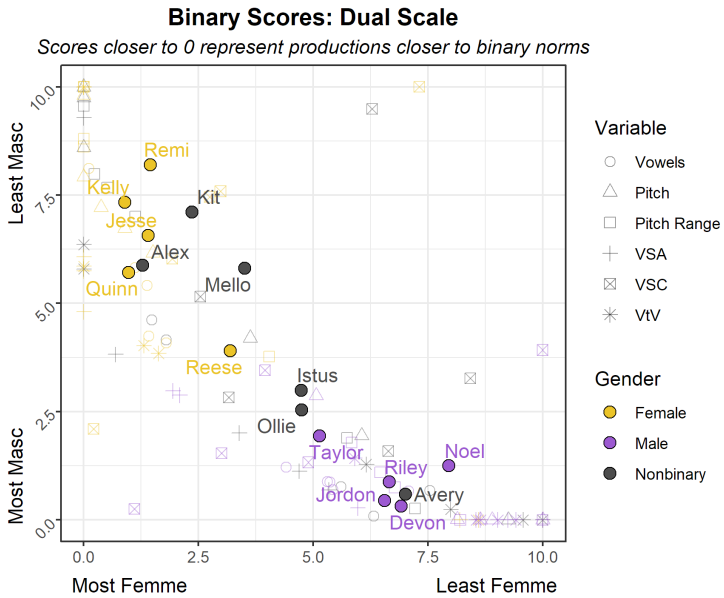


FIGURE 9.8: All variables for all speakers, plotted along dual scale. Values closer to 0 are stereotypically more masculine while values closer to 10 are stereotypically more feminine.

Figure 9.8 shows the scores used for each variable for each participant along the dual scale, whereby each production from each participant was scaled based on its deviation from the binary male and binary female means. Because these scales measure "distance from" the mean, lower scores are considered more prototypical of binary productions, with any production "equal or more gendered" than the mean being assigned a score of 0. From a broad perspective, with very few exceptions (mostly involving vowel space conformity), the results confirm the exclusive and bidirectional nature the gender binary in production and perception *for binary speakers*

– productions (and overall means) fall along a negative slope, whereby the most femme productions are also the least masc. We also see, once again, that masculine productions are more cramped, with binary men overall appearing closer to each other compared to binary women. In particular, women's productions, while relatively narrow in terms of their relationship to feminine means, are much more widely distributed when examined in relation to masculine means, compared to the binary men's productions in relation to feminine means. This is also represented in the binary women having overall larger standard deviations, especially in relation to "same gender" means. For binary speakers, the dual scale provides little insight that is not already encompassed by examining the single scale.

It is interesting that both gender groups contain obvious outliers in the form of Taylor (a binary male) and Reese (a binary female). Given the small sample size, it is impossible to know just how extreme a position these two speakers take. Certainly, from an impressionistic standpoint, I did not find anything notable or surprising in their voices during the interview or when transcribing their recordings.[5] As discussed above, Taylor himself defines his speech as firmly masculine, supporting my impression. It is feasible that Taylor and Reese are both relatively average speakers, and that another random sample would find speakers with even more cardinal productions towards both the central and extreme production space. It is also feasible that the relatively outlying features of Reese and Taylor's voices are not so outlying when other aspects of their presentation, such as clothing and the semantic content of their speech, are accounted for.

When examining the nonbinary speakers, however, the dual scale provides more nuance to an analysis than the single scale. Alex and Mello differ much more in their relationship to feminine norms than masculine norms. Mello and Kit are similar to the binary women as a group (sans Reese) with regards to masculine norms, but their productions are less feminine. Using the dual scale, both Istus and Ollie's trend toward the middle ground between groups is more obvious. I think it is useful here to point out that Istus and Ollie were assigned different sexes at birth, not because I suddenly wish to center Assigned Sex at Birth (ASAB) as a primary category, nor discard its influence wholesale, but to emphasise that its utility as a metric is tempered greatly by each speaker's individual condition and speech goals.

Separating each speaker in Figure 9.8 to plot their variables independently along the dual scale results in Figure 9.9. The men's productions are once again relatively concentrated into one area, with no productions crossing the half-way point along the y-axis (indicating less masculine productions) and very few crossing that point on the x-axis (indicting more feminine productions). This is especially true when discounting the relatively extreme positions of many speakers' measure of vowel space conformity, indicated by the crossed square. Mello once again shows the distribution of multiple signals hypothesised by **H1c** regarding incongruence via dissonance across multiple signals. Istus and Kit also show a relatively distinct separation of variables into clusters. While Istus and Ollie looked very similar when examining their variables along a single scale, as well as in considering the mean score alone along the dual scale, another story emerges here: Istus has two relatively distinct clusters of variables, while Ollie's productions form a wide, centralised distribution.

Figure 9.10 shows Ollie and Kit plotted against the mean binary scores. This figure further illustrates the way different strategies can both achieve incongruence. Ollie and Kit are the two nonbinary speakers with the largest standard deviations between

[5]Though my own experiences with Lesbian, Gay, Bi, Trans, and Others (LGBT+) people may obviously mean I am less discriminatory when it comes to "normative" speech.

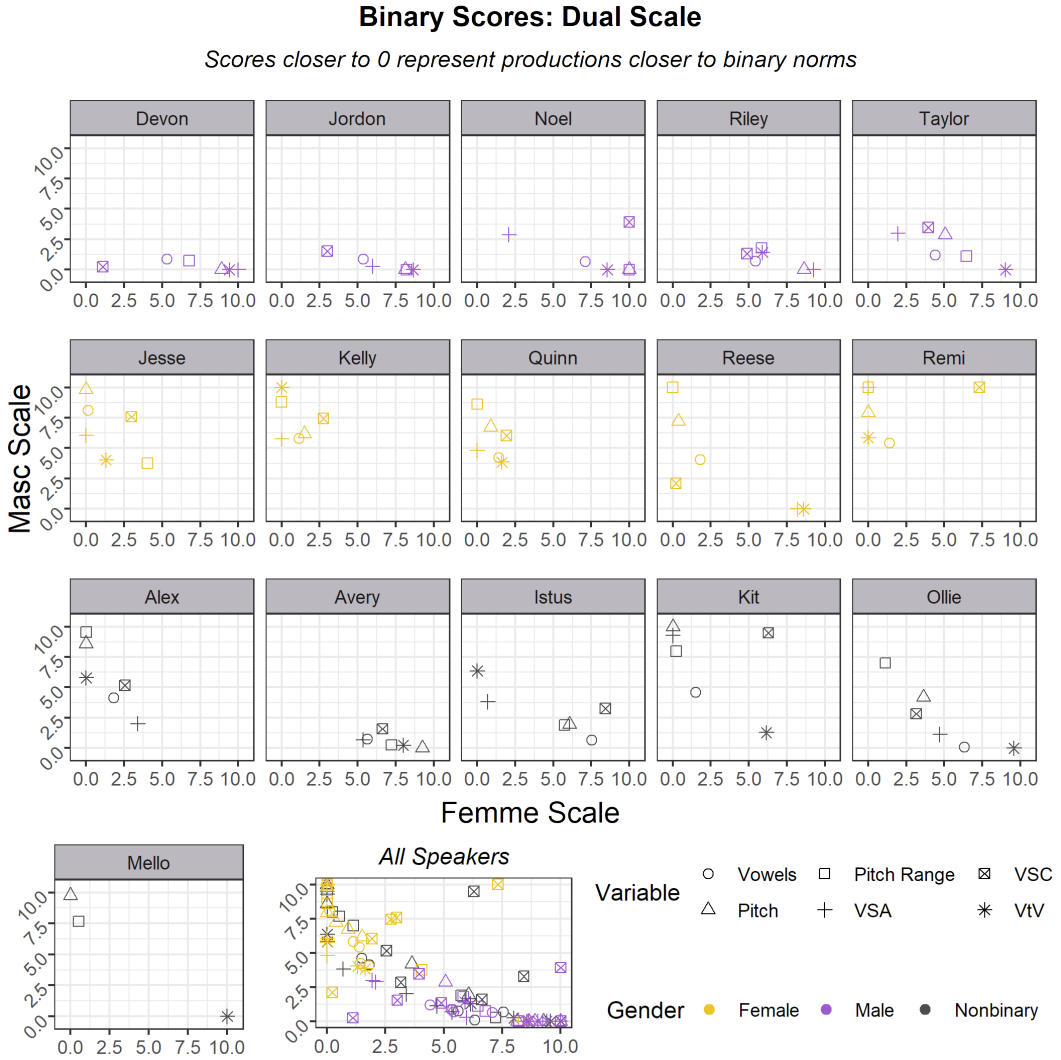


FIGURE 9.9: All variables, plotted independently for each speaker along dual scale. Values closer to 0 are stereotypically more masculines while values closer to 10 are stereotypically more feminine.

signals across the complete set of data. While both speakers' productions contain a similar spread with a standard deviation around 3 points across variables, this manifests in very different ways for each of them. Ollie's productions range between the normative means, with many tokens relatively centralised. Kit's productions, on the other hand, occupy relatively extreme positions.

9.6 Spotlights

9.6.1 Spotlight: Avery

Throughout the pairwise comparisons, Avery's speech patterns consistently with the binary men. While this runs counter to the **H1c**, Avery's productions are consistent with their identity as genderfluid and masc-aligned:

> When I'm comfortable I would say genderfluid, I feel like I probably lean towards more the masculine side but I would still identify as genderfluid.
>
> — *Interview 39:55*

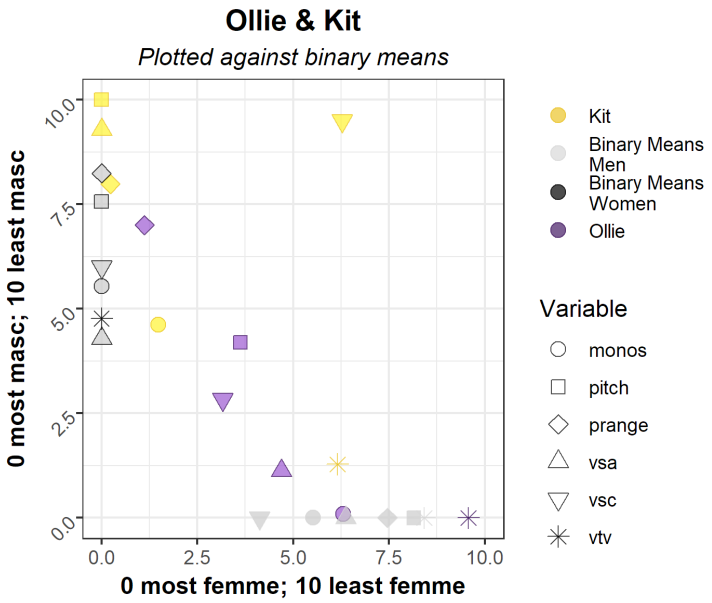


FIGURE 9.10: Multiple Signals: Ollie and Kit plotted against binary gender means. Ollie's values all trend towards a central distribution, while Kit's are widely spread out, even leaving aside Kit's extreme VSC value.

> I feel like as I've gotten older I've kind of become more comfortable with the masculine side of things. I really wouldn't want to be any different than the way I am, I don't think. But I... because I am more comfortable in that side I feel like it's also made me more comfortable just in any form of expression.
>
> — *Interview 43:23*

While Avery rejects the label "man" and the stereotypical masculinity the term represents, they embrace other masc-aligned identities like "boy" and "dude". Avery's recognition of their genderfluidity came from an intense self-reflection of their relationship to masculinity and the dysphoria that some masculine features, particularly body hair, gave them:

> I just kept looking at myself and seeing more masculine traits [...] I kind of just had this moment of, like, *I'm not happy with this* and then I kind of started thinking about it more and well... this, *you're a guy, this is what guys look like, why are you upset?* And so I started thinking about it more intensely at that point, which was probably around the same time that I started looking at alternative forms of gender. [...] Since I came to the conclusion [that I'm] genderfluid, I feel like it's become a lot easier for me to just think about it without thinking it's a big deal.
>
> — *Interview 44:38*

The realisation that they were not cis allowed Avery to manage their feelings of dysphoria. Avery's acknowledgement of their genderfluidity lifted some of the pressure they were experiencing, about the way their feelings should conform to

the expected social ideologies about cis manhood and appearance[6]. Without feeling forced to feel or act a certain way, Avery can now embrace the aspects of masculinity they resonate with, and reject the ones that cause them discomfort, without feeling like their identity is being compromised: Rejecting a cis (binary) male identity, and realising there were options outside of binary gender, gave Avery more space for, and comfort with, masculine expression.

> When I'm having a day or moment where I feel more leaning towards one or the other I would probably have that show through more in my voice, but I don't think I ever actively try to force that.
>
> — *Interview 1:08:49*

Avery's speech patterns align with that of the binary men, directly refuting **H1c**. While this is consistent with their identity, presentation, and stated speech goals (per **H3**), it is unaccounted for in a hypothesis predicting incongruence measuring only speech variables. My personal impression is that this incongruence comes across more strongly non-acoustically and in person, both in Avery's appearance, particularly their use of makeup and nail polish, and in other aspects of their communication, such as gesture and word choice.

Avery's spotlight also highlights the complexity involved in presenting a non-binary gender within the confines of the assumed binary. The idea that nonbinary gender inherently rejects binary masculinity and binary femininity equally or completely is similar to assuming binary gender is complete and proportional: it is a reduction of complexity that is ultimately dishonest, even if it may tell part of, or even most of, a story.

9.6.2 Spotlight: Istus' Frication in Context: Make-up

I return to Istus again to examine her use of frication across contexts, with particular focus on her use of makeup. While RAINBO contains selfies from each recording context, as well as many questions in the interview about each participant's use of fashion in presenting their gender, a truly thorough examination of the visual domain in identity construction is outside the ambit of this thesis, where the focus is on speech. This spotlight serves as a glimpse into just how much complexity **H1c** may imply when multiple domains are incorporated.

Figure 9.11 lists Istus' frication rates across the interview and In the Wild (ItW) recordings. In addition to the percentage of coded tokens of intervocalic /t/ produced with frication, I have also listed the number of total tokens, and a basic label to

[6] I think there is also influence here regarding *homonormative* assumptions, in that Avery's identity is strongly tied to their relationship with gay masculinity, which itself is already somewhat excluded from (and antagonised by) the heteronormative status quo.

[7] While RAINBO contains Istus' selfies from these recordings, and I believe the difference in makeup, particularly lipstick, is striking, I have decided not to include them here. When Istus agreed that her anonymised photographs would be included in the research data and presented, the implication was of a larger overall study examining the relationship between gendered presentation in the audiovisual domain. This is still the intention of the RAINBO corpus, but the primary focus of this thesis is on the auditory domain only. Transfeminine people, compared to other trans identities, are subject to hypervisibility in popular conception (Zimman, 2013, p. 34), specifically an over-representation of particular, often racialised, stereotypes of transfemininity. It is often the case that transfemme people's personhood is reduced to villain or victim (or both) and the use of their bodies as objects of scrutiny and targets of violence is normalised (Johnson and Boylorn, 2015; Berberick et al., 2018; Carroll, 2020; Jones, 2020; Hamley et al., 2021). Given that context, and the final shape of this thesis, I have chosen not to include Istus' photographs here.

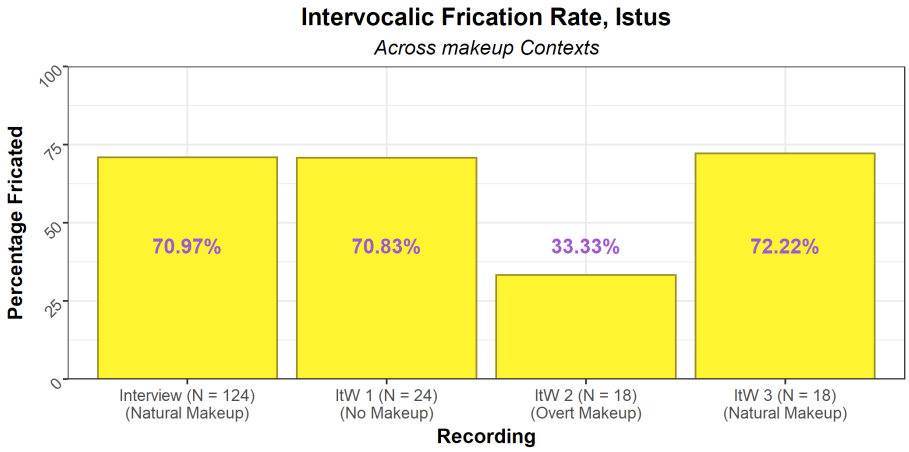


FIGURE 9.11: Istus' frication rates across recording (and makeup context).[7]

describe the kind of makeup Istus is wearing in each recording. While the overall number of tokens is very low, there is a marked difference in the rate of frication in ItW 2 compared to the others, especially considering the consistency across the other contexts.

This is what Istus says about her own use of makeup in ItW 2:

> "I'm wearing some just basic makeup stuff, but it's makeup stuff that looks like I'm wearing makeup, as opposed to what I'm wearing now."
>
> "So like I was wearing makeup for the sake of wearing makeup [...] if I wear enough gold eyeshadow it covers the bags under my eyes."
>
> — *Istus, Interview, 25:31; 29:53.*

The "now" refers to what Istus' makeup is like during the interview, which I have labelled as "natural" in the graph. While she did not explicitly describe the makeup she was wearing the day of the interview, it is similar to what she wore in ItW 3, which she describes as "very light".[8] I have labelled it as "natural makeup" to contrast with the "overt makeup" Istus describes. The context between all three ItW recordings is incredibly similar,[9] involving Istus telling a story to one other person about a particular project she has recently undertaken. Her ItW recordings are certainly much more consistent contextually than recordings provided by most other participants, and with nothing obviously different about ItW 2 in terms of interlocutor or topic that would account for the variation. Notably, her interview data shows frication rates in line with ItW recordings that match her use of make-up, though this recording is contextually much different, containing a lot more interaction between Istus and the interviewer (me), and much different topics and content. Given Istus' control of her speech and overall presentation, her awareness of the kinds of femininity she does (and does not) embrace, and her consistency in production in other contexts such as pitch, I offer the following two possible interpretations for her variation across makeup contexts:

i. Istus' use of a marker like frication may vary with her use of makeup because Istus has a "target" gendered presentation that she is aiming for overall, such

[8] "I'm wearing very light touches of makeup, my eyebrows are not that dark normally." (Interview, 45:38)

[9] All recording contexts for all speakers can be found in Appendix C.

that the "increase" in perceived femininity presented by the overt makeup (red lipstick, heavy black eyeliner) is countered by a reduction in other feminine markers, like frication.

ii. Alternatively, it could be that the type of femininity encoded by the overt makeup, and the overall combination of features that comes along with it, are different enough from the features encoded by the frication so as to be incompatible. In New Zealand English (NZE) the fricated variants are associated not just with binary women, but in particular older women of a higher social status. The use of bright red lipstick has long been associated with a type of youthful femininity that is overtly sexualised and rebellious (Jackson and Vares, 2015; Gurrieri and Drenten, 2021), an image that is overall at odds with the presumed sensibilities of the type of "proper", prestigious women for whom fricated /t/ is a marker in NZE.

Counter-evidence to these proposals is that other aspects of Istus' voice, such as pitch, are remarkably consistent throughout all her recordings, including ItW 2, when we might expect it to vary along with other acoustic markers like intervocalic /t/ frication. I do not have a satisfactory answer to that, and it is unlikely that I happened to find the one acoustic variable that is intrinsically tied to Istus' use of makeup. However, I do believe, given Istus' evident control in other aspects of her voice, the centrality of gendered presentation (in particular femininity) to her identity, and her remarkable stability across the other recordings regarding frication, that her drop in frication in ItW 2 is not random, and that the gendered perception of this feature is the driving force behind it.

9.7 Discussion

It is immediately apparent that VSC is a relatively poor indicator of gendered conformity when applied in the way I have, as it provided some inexplicable extreme scores for some participants in regards to either scale. Kit's and Remi's VSC were found to deviate remarkably from both masculine and feminine norms, with scores on both scales being higher than 6. Devon and Reese, conversely, were both found to have scores under 2.5, indicating a VSC that conformed to both binary norms. While not inherently contradictory, these results were not necessarily in line with any participant's scores along other measures. Because of the way the dual scale inflated scores for "deviance" away from gendered norms in one direction while erasing scores for conformity to 0, this kind of doubling may happen if some of the vowels forming the convex hull are close to the means for different binary norms. However, what is likely to have happened in this case is a product of the way the vowel space was calculated: Since each participant's actual productions are used to determine the outermost vowels that make up their hull, some participant's vowel spaces are calculated using different vowels than the ones that resulted from taking the means from all participants. This was appropriate in the comparative visualisations in Section 7.3.4, which overlaid the hulls while directly comparing the vowels that engendered them. However, in the VSC analysis this may cause an extreme deviation from the mean hulls if the participant's vowel productions result in a hull with a remarkably different shape.[10] While this was all intended behavior when I decided to use VSC as a form of numericising vowel dispersion via shape, the complex polygons involved and the

[10]Some of the variation is also definitely due to the values not being normalised first, but that runs into all the problems inherent with normalisation discussed in chapters 2.2.2 and 7.3.4.

limited number of speakers effecting the resulting binary "norm" hulls ultimately require a much more detailed analysis of the shape conformity than I am able to provide, and is therefore an inappropriate measure in this context.

Throughout this thesis, but particularly in examining multiple signals along the dual scale, the awareness and ability of nonbinary speakers to incorporate multiple gendered norms simultaneously becomes apparent. I have mentioned before that the binary men's productions tend to be more constrained, and 9.9 provides some insight into why this may be the case. While the productions of binary men (and their means) do tend to fall into the lower right quadrant (the production space representing both the "most masculine" and "least feminine"), the range of the men's productions is much more constrained in reference to the masculine means than the feminine ones. Only a fraction of the total productions reach a score higher than 2.5 along the 10-point scale, with none scoring higher than 4. While scores along the other axis tend to be higher in general (indicating less feminine productions), they still range more widely across the possible production space. For example, Devon's signals, according to the dual scale in Figure 9.9, range from 1.1 to 10 across the feminine axis,[11] indicating productions that deviate from feminine norms to varying degrees. Devon's productions along the masc scale, however, range from 0 to 0.88. No matter how feminine Devon's productions, no matter where they fall in relation to femme norms, they are *always* masculine. A similar pattern in the opposite direction is found for the binary women (in that productions are relatively constrained across the x-axis compared to the y-axis). This indicates that, for both binary men and binary women, conformation to "same gender" expectations is more important than deviation from "other gender" expectations. That is, it is more important for binary women to sound like women, and binary men to sound like men, than it is for either of them to *avoid sounding like the other binary gender*.

Ultimately, given the social foundation that assumes that gender is exclusive and binary (anything not male is *de facto* female and vice versa), this nuance is relatively inconsequential when examining only binary speakers, in the sense that the single and dual scale tends to place binary speakers in the same relative positions and create a clearly defined division across groups. However, the dual scale allows us to see that nonbinary speakers appear to be navigating the gendered worlds of "masculine production" and "feminine production" differently. While this is tempered by my methodological decision to have the binary speakers' data establish the norms against which everyone is measured, I would argue that this is similar to the way these norms are established and reinforced by binary speakers in the context of the shared social foundation.

9.8 Conclusion

This section has provided evidence that some, but not all, nonbinary speakers use the combination of multiple signals in constructing a nonbinary gender identity. For some speakers, like Mello and Kit, this is consistent with the hypothesised incongruence via dissonance. For other nonbinary speakers, like Istus, it is consistent with incongruence via ambiguity within single signals. Overall, this is a very promising result, considering that dissonance across signals cannot hold true at the same time as dissonance within signals.

Binary speakers, on the other hand, tend to produce multiple signals that are consistent with binary norms, though again there are some exceptions, like Reese and

[11] Or 5.3 to 10 if you remove the VSC score as an outlier.

Taylor. It has also been shown that the method of gendering and scaling responses plays a role, as what appears to be ambiguous or polar productions on a single gendered scale may represent adoption (or rejection) of femininity and masculinity differently. When examined along a dual scale whereby feminine and masculine norms are scaled separately, it is generally the case that binary speakers' productions are influenced more strongly by the norms that match their own gender, while nonbinary speakers incorporate both binary norms to varying degrees.

The spotlights provided some insights and shortcomings in applying a necessarily holistic hypothesis like **H1c** to finite data limited to one modality: While it is impossible to experimentally measure an entire person, nor is it likely that plucking any random set of variables to study will result in perfectly identifiable incongruence, the results of this analysis still lend support to the concept of multiple signal incongruence via dissonance. Nonbinary speakers are absolutely able to employ this particular form of stylistic bricolage, and manipulate their productions in response to binary norms. In particular, being nonbinary seems to allow them a greater nuance in interpreting binary stereotypes separately, and allow a wider range of expression.

This hypothesis is complex, and addressing it requires some subjective decision-making regarding the categorisation, analysis, and normalisation of multiple variables. Simple categorisation is one hurdle, of determining what makes a signal masculine, or feminine. I addressed this in the analysis by using two different approaches, one (as represented in the hypothesis) which presumes a single distribution of productions along binary lines, and a second that interprets productions based on their similarity to binary gendered norms separately. Another source of complication is in selection: Which variables to select, and where do you stop? In the initial planning and creation of the RAINBO corpus, I asked participants to take selfies to accompany each recording, and asked a number of questions regarding their mood and physical presentation. As discussed in Chapter 4, every interview in RAINBO is also fully demarcated by topic using three different critera (topic, tense/aspect, and theme). The intention when collecting this information was to help address **H1c** (and **H2**), and to examine the ways that language and the voice interact with non-acoustic aspects of presentation in forming identity. It is well-established that physical appearance and the perception of self has major consequences on speech production and perception (Strand, 1999; Van Borsel, De Cuypere, and Berghe, 2001; Van Borsel, Pot, and De Cuypere, 2009), and I had hoped to examine the creation of incongruence via cross-modal signals. Due to time constraints, I was not able to adequately address this data (except in Section 9.6 where I briefly discussed Istus' use of makeup in tandem with her frication), but this data exists in RAINBO for future analysis, and I am sure it would add a fruitful and nuanced angle to the discussion.

Chapter 10

Discussion & Conclusion

10.1 Introduction

This purpose of this thesis has been to examine the ways nonbinary people enact identity through their speech, and to situate an examination of nonbinary speech alongside established explorations of binary-gendered speakers. It has done so via the creation of the Recorded Audiovisual Interviews with Nonbinary and Binary Orators corpus, a number of qualitative and quantitative examinations of phonetic variation in single variables across populations within RAINBO's collection, and finally an analysis of multiple signals together.

In the introduction to this work, I started with the broad research questions, then the particular hypotheses being tested, and finally the methodology and analysis to answer those research questions. In this final chapter, I take the reverse approach. I begin with a brief summary of each analysis in Section 10.2. In Section 10.3 I engage each hypothesis in turn, then address the broader research questions in Section 10.4. I contextualise the work itself by outlining several limitations and contributions (Section 10.4.2), providing a reflexive analysis of the methodology, and offering some potential directions for future work. I conclude in Section 10.5.

10.2 Summary of Work

Chapters 1 through 3 introduced the nonbinary population, the overall research questions and specific hypotheses regarding nonbinary speakers' use of speech in presenting their identity, as well as a sketch of the current and background literature that the research questions build upon. Chapter 4 outlined the research context, participant recruitment, and the creation of RAINBO. Chapter 5 introduced each of the 16 participants "in their own words" as well as my impressions of them as a researcher. Chapter 6 used two different methods of measuring pitch (Praat and Talkin (2015)'s **R**obust **E**poch **A**nd **P**itch **E**stimato**R**. (REAPER)) to analyze pitch and pitch range, and found intriguing differences across speaker, context, and gender groups in the number of modes engaged. Chapter 7 examined monophthongs, the overall Vowel Space Area (VSA) of speakers, and the distribution of NEAR/SQUARE vowels among the older triad. Chapter 8 examined intervocalic /t/ frication, a categorical signal that appears to encode a nuanced femininity. Chapter 9 presented an analysis of the aggregated results from the previous chapters, to examine the combined effect of multiple signals on gendered production, and found evidence that binary speakers treat binary gender as a single scale while nonbinary speakers respond differently to the social encoding of masculinity and femininity in speech.

10.3 Addressing the Hypotheses

In this section I address each hypothesis separately. After reiterating each hypothesis, I outline the results from each analysis in terms of whether they provide confirmatory, contrary, or inconclusive evidence of each prediction.

10.3.1 H1 - Acoustic Incongruence

Hypothesis H1a – Categorical Signal Incongruence – *Given a single categorical variable, nonbinary speakers can communicate identity via **dissonance**, by producing tokens at rates between those found for binary male and binary female speakers, or via **ambiguity**, by producing a variant not found in either men's or women's productions.*

H1a is inconclusive due to lack of data. The only categorical variable examined was the frication of intervocalic /t/, which yielded only 2 categories once the data was cleaned, with mixed results. As a group, the nonbinary speakers were not found to produce rates of frication significantly differently from the binary men or women, though men and women produced frication significantly differently from each other. Among the older triad, the nonbinary speaker, Alex, produced rates of frication consistent with incongruence via dissonance. However, among the younger speakers of all genders, the data suggest the interpretation of intervocalic frication is not just feminine, but encodes a particular form of femininity that not all femme-aligned speakers ascribe to at all points in time. A qualitative examination of two nonbinary speakers, Istus and Kit, highlighted this finding, as they both showed variability in frication according to different contextual factors: Istus in regards to her positioning herself in the discourse as "outside" femininity looking in, and Kit in regards to their emotional state and the interlocutors present evoking social fatigue. More data, from more speakers, from different categorical variables, are needed to accurately examine the avenue of incongruence hypothesised in **H1a**, but these initial findings are intriguing.

H1b: Gradient Signal Incongruence

Hypothesis H1b – Gradient Signal Incongruence – *Given a single gradient variable, nonbinary speakers can communicate identity by producing incongruence either via **ambiguity** (in the form of more centrally-distributed tokens compared to binary male and binary female speakers), or via **dissonance** (by producing tokens in a bimodal fashion, where a single speaker's productions are within both binary male and binary female ranges).*

Table 10.1 collects the results from multiple analyses of single signals, pointing out where the data from individual speakers (or gender groups) showed evidence supporting and refuting **H1b**, including the type of incongruence created, as well as complicating factors. Along with each result, I have added an overall judgement of whether the hypotheses are "supported", "refuted", or "inconclusive", based on my impression of the overall strength of the evidence. (For example, while the results of the pitch range analysis found a lot of evidence for the hypothesized forms of incongruence among nonbinary speakers, the counter-evidence from the vowel variation within all gender groups and the complicating finding that women's pitch productions aligned with the predictions for incongruence via dissonance, lead to an overall inconclusive judgement.) Due to the lack of templates and the position of nonbinary speakers as pioneers in creating a "nonbinary voice", the hypotheses proposed were never intended to hold true in all cases at all times for all speakers: A speaker producing incongruence via dissonance cannot also produce it via ambiguity

Variable	Hypothesis	Type of Incongruence	*Group* or Individual	Notes & Complications	Conclusion
Pitch Means	**H1b**	Ambiguity	Istus		Support
			Alex		Support
		Dissonance	Kit		Support
			Istus	Utilised as a joke.	Inconclusive
	Complications			*n.a.*	
					Support
Pitch Range	**H1b**	Ambiguity	Istus		Support
			Ollie		Support
		Dissonance	Alex		Support
			Nonbinary	Multiple pitch modes per speaker.	Support
	Complications		*Women*	Most women *also* produced multiple pitch modes.	Inconclusive
			All	Pitch range wider within- than across-group.	Refuted
					Inconclusive
Normalised Monophthongs	**H1b**	Ambiguity	*Nonbinary*		Support
	Complications		*All*	Differences are not statistically significant across groups.	Inconclusive
				Not enough data.	Inconclusive
					Inconclusive
Unnormalised Monophthongs	**H1b**	Ambiguity	Alex		Support
			Nonbinary	F2 Significantly lower than women, n.s.d. with men.	Inconclusive
		Dissonance	Kit		Support
	Complications		*All*	Not enough data.	Inconclusive
			Istus	Productions outside of predicted values for both binary norms and incongruence.	Inconclusive
					Inconclusive
VSA	**H1b**	Ambiguity	Alex		Support
		Dissonance			*n.a.*
	Complications		*All*	Not enough data. No binary differentiation	Inconclusive
					Inconclusive
VSC	**H1b**	Ambiguity	Alex		Support
		Dissonance	Kit		Support
	Complications		Istus	Productions outside of predicted values for both binary norms and incongruence	Inconclusive
					Support

TABLE 10.1: Results of all analyses testing **H1b**, indicating evidence that supports (or refutes) the predictions made for creating incongruence via ambiguity and incongruence via dissonance, including complicating factors.

within the same signal, and a speaker creating dissonance across multiple signals as proposed in **H1c** cannot also create it within individual signals as proposed in **H1b**. It is unsurprising, therefore, that the results are complex and the higher-level group statistics tend not to reach a level of suggesting significant difference.

Support for both forms of incongruence were found among nonbinary speakers. Where a nonbinary speaker's productions in one variable were found to align with the predictions for creating incongruence, it tended to be the case that the same form on incongruent production was identified in other variables. However, at least one binary speaker, Avery, provided a strong refutation of the creation of incongruence, as their productions for all examined variables were within the masculine production ranges. The overall lack of data and insufficient differentiation found in binary populations means support for this hypothesis rests on relatively weak foundations: The hypothesis assumes a robust social foundation of binary stereotypes for nonbinary speakers to work with, and while these stereotypes are often well attested in other literature, they were not necessarily supported within the demographically-matched binary population in RAINBO. More evidence from a greater population of speakers, or a much closer examination of speakers in this and other New Zealand English (NZE) corpora for evidence of incongruence, is warranted.

These results suggest the creation of incongruence within signals may be one way, of many, open to nonbinary speakers to present identity. Overall, where incongruence is found in a given speaker, it is consistent across variables. The exception here is Alex, whose productions align with incongruence via ambiguity in vowel production, but whose pitch ranges suggest dissonance. However, this is complicated somewhat by the multimodal pitch ranges of most binary women: Ambiguity via dissonance in pitch range was predicted to manifest as a multimodal pitch distribution, with either mode appearing within the production space expected of of masculine and feminine norms, respectively. Alex's productions firmly meet this criteria. However, the binary women are also *all* producing multimodal pitch ranges, which was not accounted for in the original assumptions driving the identification of incongruence via dissonance.

H1c: Multiple Signal Incongruence via Dissonance

Hypothesis H1c – Multiple Signal Dissonance – *Given multiple signals, nonbinary speakers can communicate identity by producing incongruence* ***across*** *signals.*

Evidence for **H1c** was found among three of the nonbinary speakers: Ollie; Mello; and Kit. Two of the other nonbinary speakers, Istus and Alex, presented incongruence via ambiguity in multiple single signals, but this was not sustained reliably across all signals. One nonbinary speaker, Avery, directly refutes **H1c**, and the overall incongruence hypothesis, as all of their productions are firmly in line with masculine norms. The examination of multiple signals also highlighted the ways that nonbinary speakers may be (re)interpreting gendered cues: Binary speakers tended to achieve similar scores regardless of whether productions were measured along separate masculine/feminine scores, or a single collapsed gender score, while nonbinary participants' scores varied in reference to different binary genders.

As an analysis of multiple signals relies heavily on the integrity of the individual signal data that composes it, the shortcomings identified regarding **H1a** and **H1b** likewise apply to **H1c** and make a clear confirmation or refutation ill-advised. However, in examining the speech of these six nonbinary individuals, it is clear that different production strategies are being utilised and that these strategies result in a measurable difference in speech, when compared to binary peers.

10.3.2 H2 - External Context

Hypothesis 2 – *Immediate context has a greater influence on nonbinary speakers than binary speakers*.

Overall, **H2** is refuted. In the analyses of VSA and intervocalic /t/, not enough data was available to examine the effects of context on production. However, in the analyses of both pitch and monophthongs, binary-gender groups, particularly women, were found to exhibit much higher variation within groups than across them. Istus, a nonbinary participant, maintained the highest consistency in production across multiple variables, while Avery, another nonbinary participant, produced patterns consistent with binary men.

This may be a temporal/trajectory matter. The crux of **H2** relies on an argument that nonbinary identity is a relatively new label for participants. This is contrasted with binary participants who have, presumably, associated themselves with their gender for as long (or longer) than they have been able to understand the relationship between gender and Assigned Sex at Birth (ASAB). The hypothesis predicts that this adoption of a nonbinary label (or putting a name to an identity one has already felt) would come with some aspect of instability, where speakers must navigate a "new" way to interact with the world, newer than their relationships to many interlocutors. One alternative view to this is that the "nonbinary" label is a new word for an old feeling: For every nonbinary participant interviewed, the adoption of a label is a reflection of something they have always felt, or an additional nuance to an identity they have already found familiarity in. Interpreted this way, it is less obvious that speakers would be influenced greatly by context, because that context has always been there, even when it was unnamed. This is in contrast to the experiences of many binary transgender people, who may have always "felt" trans, but the pressures of "passing" in society has often meant a more forced adoption of "new" ways of speaking, acting, and doing. While I argue the lack of a template for "how to sound nonbinary" has led to strategies for incongruence (which I have shown some evidence for, for some speakers), it may also contribute to a lack of pressure for changing speech in context.

Another way the lack of support for **H2** may be a temporal issue is in the way the methodology and research context has encouraged participation from nonbinary participants who are already comfortable in their gender identity. I have examined this context in detail in Section 4.3.1, to explain why the participants within this study are more comfortable and invested in their nonbinary identity and label compared to participants who may still be exploring options or questioning (and, therefore, more likely to present greater contextual variation). Istus' productions represent one of the more robust arguments for this kind of separation, as she discusses at length the kinds of ways she approached presenting gender in her voice and how she finally settled on a "new" way to talk that best represented her gender. In this case, the lack of a template resulted not in a lack of pressure, or an increase in contextual influence, but DIY: Istus attempted many different blueprints for speech, such as her "singing voice" and her "California girl" voice, and rejected them all as not being representative of her identity. Rather than reacting to the immediate context, Istus instead invented her own template, her own blueprints of how she wanted to sound, and uses that as the baseline against contextual influence.

Finally, the paucity of the data in making it difficult to address this particular hypothesis cannot be overstated. The kinds of contexts examined by the data available are relatively limited, looking at overall pitch and vowel variation across recordings, without consideration of the specific details, such as interlocutor or topic. While there

are good reasons to reject **H2**, there are equally good reasons to explore it more deeply using other variables.

10.3.3 H3 - Individual Condition

Hypothesis 3 – *Variation is bound by the individual circumstances of each participant. Individuals with shared socially-encoded circumstances, such as age and binary gender, will speak more like each other.*

H3 made two explicit predictions:

1. Due to greater similarity in the individual condition, less variation will be found within individuals who share a binary gender than within nonbinary individuals.

2. Variation within the nonbinary gender group will conform to each speaker's individual condition, in terms of identity, history, and intention.

Regarding 1, the data from the analysis of VSA and intervocalic /t/ show wide variation, found within all examined groups, regardless of gender. However, support is found in the analyses of pitch range and multiple signals, both of which found greater consistency within binary-gender groups than across gender groups or in comparison to the nonbinary productions. Overall, this represents relatively weak support for the prediction that binary gender leads to a greater similarity in individual condition.

While outside of the overall scope of this thesis, which focuses on nonbinary speakers, an interesting aspect of **H3** arises when examining the binary gender responses more carefully. Overall, across variables, the binary men are found to exhibit the highest within-group conformity. This can be seen most clearly in the previous chapter on multiple signals, Chapter 9, in the way male productions cluster much more closely together than women or nonbinary participants. This also leads to a further question: Is the conformity found within-group a product of the individual condition, as hypothesised, or is it a product of the pressure exerted by the shared social foundation? The results here suggest it might be both, or that the influence of the latter should not be underestimated when examining gender in linguistics (whether that work incorporates nonbinary speakers or not). Measurements of male speech are often taken as the default, but the evidence here suggests that men are subject to a possible restriction of their "natural" variation.

Regarding 2, there is substantial support for the influence of the individual condition on nonbinary productions. In the preceding examinations of pitch, vowels, frication, and multiple signals together, I have presented spotlight data exploring the relationship between the gender of individual speakers and speech production.

Ultimately, it is unsurprising **H3** found robust support within this analysis, as it forms the scaffolding for **H1**, and the reason that different manifestations of incongruence are predicted for different speakers. Examples like Avery's productions aligning with masculine norms is not necessarily a bug of **H1c**, but a feature of **H3**.

10.3.4 Summary

Some support was found for the creation of incongruence (**H1**) in presenting nonbinary identity, among some speakers, in both categorical (**H1a**) and gradient signals (**H1b**), as well as across multiple signals examined together (**H1c**). However, addressing **H1** and **H2**, particularly the predictions encompassed in **H2**, was complicated

due to the overall lack of data. In many cases, the numbers of participants or data points were just not high enough for a reliable statistical analysis. Further, this lack of data made it difficult to establish the binary baselines required to establish what gender incongruence would look like in a given analysis. However, this focus on a smaller number of participants and examination of individual data also led to more robust support for the influence of the individual condition predicted by **H3**, via qualitative analysis. Overall, the findings from this work support previous work on nonbinary speakers that show they are hyperaware of prevailing binary stereotypes in speech (Kirtley, 2015; Corwin, 2017), the relationship between their individual condition (such as their bodies or pitch) and the social fabric that genders those bodies in certain ways (Zimman, 2014). It also found, as in previous work, that some nonbinary speakers discursively construct their identity through their conception of their relationship to these binary stereotypes (such as Istus' California voice, or Avery's amnesty of masculinity) (Cordoba, 2020; Garmpi, 2021). It extended both of these findings from that found in previous work by incorporating an examination of binary speakers (finding some nuances in gendered production, such as in pitch range and /t/ frication), older speakers, and a combined qualitative and quantitative approach that specifically focuses on the individual condition to inform and support the results of statistical analyses.

10.4 Addressing the Research Question

I now step back from the detailed hypotheses, and address the research questions that have guided this dissertation:

1. Given their status as an emergent, distributed, and minority population, does nonbinary speakers' use of gendered speech features differ from that of binary speakers due to their nonbinary gender?

2. Given the pervasive historic and enforced gender binary, to what extent do existing methods of phonetic analysis support the inclusion of nonbinary speakers?

Regarding 1, I have demonstrated through both qualitative and quantitative analysis that this is the case for some nonbinary speakers, though some caveats are required. The first is that I *also* found evidence that at least one nonbinary speaker, Avery, did not appear to alter their speech towards an incongruent presentation, or at least did not alter the variables studied, on the basis of their nonbinary gender. This may be expected precisely because of the emergent, distributed aspect of a nonbinary dialect in formation. I have discussed in Chapter 9 that Avery's identification as nonbinary allowed them to be more comfortable with their own relationship to masculinity, which may have allowed them to embrace a form of speaking more in line with binary norms. The second caveat regards the causal link proposed between variability and nonbinary gender. The current data cannot establish in most cases whether measured variation is due to speaker sex/gender, not just in distinguishing nonbinary speech but also in distinguishing binary gender groups from each other. Where variation along gendered features has been established, there was also variation within same-gender binary speakers, challenging the causal link between nonbinary speech and the way binary speakers actually speak. This may indicate that speakers respond to speech stereotypes over and beyond the speech they actually encounter from binary peers. It is also possible the variation found

is not a product of sex/gender, but another aspect of identity or personality I did not test for, or even a combination of sex/gender and other features (such as the particular forms of femininity encoded by intervocalic /t/). However, the qualitative data, particularly from Istus, present a strong argument that at least some variation in nonbinary speakers is a direct result of their nonbinary gender, and reflects these speakers' complicated relationship to the sex/gender binary.

Regarding 2, I believe I have sufficiently shown that certain traditional methodologies in sociophonetics need to be approached with care and revised when attempting to incorporate nonbinary speakers. While somewhat unexpected, the largest of these issues is before the recording equipment even turns on. My recruitment notices, both the physical posters and on Facebook, turned into an obvious source of aggression, both towards my participants, myself, and for bystanders. The toilets nearest the interviewing space were likewise a site of public discourse and potential distress to nonbinary participants, while also a potential source of security, as they eventually became intentionally all-gender. While none of these issues were completely in my control as a researcher, and I sought to mitigate the issues as best I could after becoming aware of them (for example, rewording and more closely moderating electronic recruitment notices, placing physical notices closer to areas within sight of administrative staff[1], and proactively letting participants know about the toilets before the interview), this still created a more stressful recruitment context for nonbinary participants compared to binary participants, and I think this awareness (and disclosure) of context in future work recruiting marginalised populations is a necessary consideration.

Regarding the write-in demographic forms (which can be found in Appendix B), while there was minimal variation among the ten binary participants, there was still some variation (for example, whether someone proactively identified themselves as cis, or listed their sexuality as well as their gender). I think this is a clear case where the traditional methods – that is, providing a binary checkbox of "Male or Female" – serve to erase potential sources of interesting variation. Whether someone emphasises their relationship to ASAB or attraction, when given the chance, tells us something about the way they interpret their sex/gender that may be different to other participants. Beyond that, in larger scale studies it may be fruitful to examine the specific words people use even within binaries. Participants of all genders expressed varying levels of comfort with gendered terms and being treated in gendered ways, but I provide Avery's here:

> [...][If someone says] "You've grown into a good man." or something like that, and I'll be like *I'm not a man! I'm a dude! I'm a boy! Leave me alone!* [laughs] Like for whatever reason, *boy* and *dude* just sound less gendered to me than *man* because I just, I can't view myself in that stereotypical masculinity.
>
> — Avery, Interview 07:21.

Moving beyond recruitment methodologies and to other forms of analysis, the analysis of pitch range suggested that improvements could be made in incorporating nonbinary gender in future work. While number of participants was small, the data suggest that stereotypes within the existing literature regarding men's and women's pitch ranges may be inaccurate, and that gendered differences found might be more complex than "women have a wider pitch range than men." The evidence from

[1]Though it should be noted that increasing the visibility of the notices also increased the visibility of potential participants, and not all administrative staff can be assumed to be safe or supportive, either.

this group of speakers is that women are more likely to be utilising multiple pitch modes, and while this indeed represents an overall wider pitch range, it is not the case that each of these modes is itself necessarily wider, nor equally utilised. The point may be relatively moot when examining binary speakers, however – After all, men and women still clearly fall into two groups, and women's upper pitch mode is usually much denser than their other modes. Attempting to incorporate and interpret data from nonbinary participants leads to a number of questions regarding the gendering of pitch. For example: In presenting feminine pitch, is the overall range more important than the existence of two or more modes? If it is the latter, is the location of the modes or the location of the antimode the feature that drives perception? Do multiple modes represent multiple "registers" in speech, do they align with certain parts of speech (such as an abundance of phrase-final uptalk), or are they more extreme highs and lows within single units (at the word or phrase level)? These are just a few possible considerations for future research on gender and pitch that does not even require special consideration for nonbinary participants.

10.4.1 Contributions

Beyond the broader research questions, this work has contributed in a number of ways to the greater body of work addressing the sociophonetics of sex/gender. The first is in the dataset itself, and the second and third regard the way the analytical framing and my methodological approach have supported both the incorporation of "new" populations in research, such as nonbinary people, as well as the way "new" approaches to established methodology may open the space for new data from "old" populations, like allowing binary participants more self-determination is naming and defining their own sex/gender.

The RAINBO corpus is unique (as far as I am aware) in that it contains audiovisual data from both nonbinary speakers and matched binary peers. This enabled the accurate comparison of nonbinary speech to the binary speech they encounter every day, rather than being forced to rely on a combination of known stereotypes and previous corpus work. Due to the data being collected in Ōtautahi Christchurch, Aotearoa New Zealand, the binary speakers may also serve as a new timepoint in examinations involving other corpora and projects in NZE (such as ONZE). The way data was collected and licensed provides a rich dataset for future research, as well as the possibility of later expansion to incorporate more speakers.

By incorporating the examination of nonbinary speakers alongside binary speakers I have attempted to address some of the critiques against assumptive binaries in analysis (Zimman, 2014). At the very least, examining binary men and binary women alongside nonbinary people, and framing research questions and analysis in a way that puts all groups on equal footing, I both reduce the "othering" of nonbinary people, and turn a mirror on the way some analyses are rooted in normative practice, but they do not have to be. An unintentional but positive consequence of this approach is that it highlighted the possibility for increasing nuance to the existing conversations regarding binary gender, among research with binary people. The research methodology was constructed particularly to excise binary assumptions (for example, all paperwork used gender-neutral terminology, I separated questions of sex/gender from those of attraction/sexuality, I asked about participant gender using fill-in-the-blank, and in the interviews asked all participants about their impressions and relationship to binary gender stereotypes without implying exclusive membership to one or the other). While the intention of this was for the comfort of

nonbinary participants, the resulting data evidences that it allowed binary participants a little more space as well, to disclose or assert other aspects of their complex identity beyond sex/gender (such as race/ethnicity, attraction, religion), and how their own conceptions of sex/gender intersected with the hegemonic pressure to conform to a gender that is not just binary, but is white, and able, and colonising, etc.[2] Any research involving human subjects, whether incorporating Lesbian, Gay, Bi, Trans, and Others (LGBT+) populations or not, would benefit from more of this kind of interrogation of baseline assumptions.

This work has contributed to the methodological conversations within sociolinguistics in a number of ways. In examining unnormalised vowel data, I have identified evidence that some nonbinary speakers affect system-wide shifts in production, indicating there is useful, meaningful sociolinguistic information in the data that is often normalised away. In comparing multiple methods for measuring pitch, I have identified that some speakers utilise multiple pitch modes in production even after accounting for creak, indicating gendered differences in pitch range may be more accurately described as differences in pitch range**s**. At a number of points I have attempted to expand the historical record to incorporate gender beyond the binary into existing analyses, particularly in examining the NEAR/SQUARE vowel merger. This dissertation is a major contribution to the existing body of work on nonbinary speakers in both corpus phonetics and mixed qualitative/quantitative sociophonetics. As far as I am aware, RAINBO is the first audio corpus to intentionally recruit both nonbinary and binary participants using the same research paradigm, and this dissertation is the first examination of the sociophonetics of nonbinary gender in NZE.

10.4.2 Limitations & Reflections

As mentioned throughout the work, while the strength of the anlysis is in its balance between larger-scale quantitative data with smaller-scale qualitative examinations, for some forms of analysis the data was both too big and too small. From a quantitative perspective, the data is too small: There are simply not enough participants, and not enough suitably processed data to work with. This is evident in the weak (or absent) power of the statistical results throughout the dissertation. While I have endeavored to make clear the nature of the analyses as a balancing act, and without equating weak evidence with *no* evidence, the lack of large statistical power when attempting statistical analysis is obvious, and annoying. Conversely, from an ethnographic and qualitative standpoint the data is too big: There are certain limitations inherent in examining so many speakers from disparate backgrounds. Other work with an ethnographic component has focused on fewer speakers (Kirtley, 2015), or less data.

There is no right answer in terms of "how deep" or "how broad" an analysis goes, and I suspect I would be equally unsatisfied with the amount of work left "on the table" had I taken a different approach. Every section that would benefit from a deeper analysis – and that is *every* section – would mean leaving another area unexplored. Any disappointment for things I have designated as out of ambit (the influence of topics or interlocutor, other diphthongs, the role of silence or pauses, speech rate, etc.) is tempered by the success of striking a balance between the big and small, and been able to use each to inform the analysis of the other in a way that would be impossible if the focus were shifted.

[2] Nonbinary participants talked about this too! But I think, for the nonbinary participants, it was much more obvious that their gender intersected with all of these other aspects of life, while this is not always the case when your research population is presumed to be part of the dominant culture.

As a nonbinary person, examining the speech of nonbinary people, every exploration feels like both expedition and self-discovery. I think everyone who engages in this kind of self-adjacent research finds an aspect of affirmation in it, as I certainly do. With that proximity there also comes a feeling that not doing it completely at every step is a somewhat personal failure. That said, I recognise that it is impossible to do everything, and that these feelings are part of the research process as well. There will always be more to explore, and the opportunity for a closer look. However, there are a number of more concrete shortcomings of this work I would like to address specifically, because they represent something beyond general dissatisfaction at not being able to do more things, better, in less time:

Accommodation

As originally conceived, I positioned myself as a researcher in a much more traditional, Labovian sense. However, shortly after beginning the interviews I became quite concerned about accommodation effects, especially as regards my (Canadian) accent positioning me as an outsider to all participants while my (nonbinary) gender (and ASAB) created a variable insider/outsider status with each participant. As well, given the assumed lack of nonbinary speech templates onto which the hypotheses rely, my position as an authority may serve to strengthen any accommodation effects. For every interview after the first (Avery), I recorded myself as well as the participants, and transcribed myself as well.

This proved to be, essentially, an unpleasant[3] timesink. When it became apparent that transcribing everything in time would be impossible, I chose to prioritise first the In the Wild (ItW) recordings, and secondly all recordings containing nonbinary speakers. In retrospect, a better approach would have been to prioritise partial transcription of all participants. The lack of binary interview data (and therefore most comparative analyses) is apparent throughout this work.

An additional prioritisation applied to the transcription. I approached the quantitative analysis with the intent of comparing speakers of different genders, and in doing so made decisions in transcription that prioritised speakers that were similar to each other. I mentioned them briefly in Chapter 5, but Flynn and Keegan were both removed from the current analysis due to "too much" variation in their speech, compared to the other speakers, due to the effect of their individual condition / language acquisition meaning they were not speakers of standard (Pākehā) NZE. The variation in their voices indicated their data would be impossible to include in any quantitative analysis, which justified their removal (especially considering the overhead involved in transcription). From a qualitative perspective, however, their inclusion in RAINBO is invaluable and a closer analysis of their speech would (and will, I hope) provide a much more nuanced take of the interaction between nonbinary gender and other aspects of identity in the future.

Intersectionality, Biculturalism and the Aotearoa New Zealand Context

Kimberlé Crenshaw (Crenshaw, 2017) first identified the ways individuals, particularly Black women, are failed by a legal system that addresses those two aspects of race and gender/sex separately, and defined intersectionality as a way of conceiving of, and accounting for, the unique situation that this multiplicity of identity requires

[3] I do not recommend transcribing yourself if it can be helped it at all, as you get to relive every faux pas or awkward phrasing, as well as gain insight into all the times you put your foot in your mouth without realising it.

in this context. The single-axis categorisation (Crenshaw, 1989) erases the unique, multidimensional experiences at the intersection of the two (or more) identities. Intersectionality in a more general, nonlegal sense, refers more broadly to a research paradigm that treats aspects of identity as in a "interactive, mutually constitutive relationship" (Hancock, 2007, p. 67), overlapping rather than discrete.

I regret that this work fails in adequately addressing intersectionality and, in particular, Aotearoa New Zealand's status as a bicultural society situated within a colonial history. There is an undeniable influence of Te Tiriti o Waitangi and the shared Māori and Pākehā history on the perception and policing of gender, and my participants' placement of themselves within that larger context. For a thesis that emphasises context, its absence is both obvious and indefensible. From early in the planning process, my research deliberately excluded speakers of Māori English.[4] I justified this exclusion on three grounds: firstly, I had been in Aotearoa New Zealand only a few months at that point, and while I was aware of the two dialects, I was not confident in my ability to identify them while interviewing or transcribing; secondly, I felt it was too difficult to try to compare multiple dialects given the small scale of the study (these are the same grounds on which I excluded Flynn and Keegan); thirdly, in my position as an outsider, I did not feel equipped to address the cultural nuance and history of takatāpui identity and how it relates to the broader rainbow whānau in Aotearoa New Zealand, nor undertake this kind of work without community collaboration. In hindsight, while I now feel more confident in my "ear" for the Englishes spoken in Aotearoa New Zealand, the other two things are still very true. However, whether true or not, none of these things are justifiable reasons to exclude participants, and my approach has perpetuated a colonial tradition of erasing and marginalising indigenous knowledge, history, and identity. My intention here is not to self-flagellate, but to take ownership of the mistake and resolve to do better in the future.

10.5 Conclusion to the Conclusion

When a child is born, the delivering doctor runs through a checklist: Ten fingers, ten toes, etcetera, etcetera. Life begins with a checklist of discrete criteria, by which our bodies are compared. Present the wrong number of toes or an irregular heartbeat, and life becomes immediately *complicated*. However, even with the right number of toes and a strongly-beating heart and a cisgender satisfaction with one's ASAB, life is not automatically simple. Complexity, for some, just takes a little longer to catch up.

By examining a population that is neither exclusively male nor exclusively female, this thesis explored a handful of phonetic features and how they contribute to the presentation of gendered language, and a gendered self outside the binary assumptions built into the social fabric. In doing so, I have highlighted the sheer amount of complexity involved in the creation and maintenance of the sex/gender binary, as well as the ability of individuals of any gender to dismantle it.

[4] I did not turn any research participant away for this reason, but the exclusion was made explicit in my ethics application and I deliberately avoided words like takatāpui in my recruitment sheets. So even if the result was the same, my intention at the time was exclusionary.

Appendix A

Glossary of Terms and Abbreviations

This glossary contains the definitions of acronyms and words used throughout the dissertation. Terms, particularly the terms regarding gender and sex, are defined as they are used specifically within the scope of this thesis.

AFAB Assigned Female at Birth

AMAB Assigned Male at Birth

Ambiguity A type of incongruence arising from the use of signals perceptually assumed to be equally applicable to all identities to which those signals might apply. For example, if the colour black is associated with one identity, and white is associated with the other identity, both grey and clear are potential forms of incongruence via ambiguity. *Gender incongruence* refers to signals assumed to apply reliably to neither binary gender. See 3.2 for an expanded definition of incongruence with specific reference to this dissertation's hypotheses.

ASAB Assigned Sex at Birth

Assigned Sex at Birth Refers to the gender/sex determination made, usually by a doctor and usually based on the infant's genital phenotype (either at birth or in anticipation of the outcome of "corrective" adjustments (Minto et al., 2003).)

Bricolage In speech, the recombination of signals with existing social meaning to create a new impression. (Eckert, 2003; Zimman, 2017).

Cisgender heterosexual (or *cishet*) Someone whose binary gender aligns with the one they were assigned at birth, who is attracted to the alternate binary gender: Cishet men are attracted to women, cishet woman are attracted to men.

Cisgender (or *cis*) Someone whose gender aligns with the one they were assigned at birth. The opposite of *trans*.

Communities of Practice The intentional grouping of speakers based on the presumption of commonality, whether real or imagined, in which members share practices and traits (such as ideals, demographics, area, attitude, etc). (King, 2019, p. 16).

Dissonance A type of incongruence arising from the combined use of signals perceptually assumed to apply exclusively to one of all identities to which those signals are appropriate. For example, if the colour black is associated with one identity, and the colour white is associated with the other identity, a single individual signaling both black and white is creating incongruence via dissonance. See 3.2 for an expanded definition of incongruence with specific reference to gender.

Enregisterment The process by which signs (including linguistics signs) become associated with (index) particular cultural practices and the types of people who engage in those practices. (Agha, 2003).

Erasure The semiotic process by which variation that violates ideological presumptions is rendered invisible, such as dialectal variation within communities often being ignored in favour of in-group/out-group distinctions. (Irvine, Gal, and Kroskrity, 2009).

Ethnography Within linguistics, the examination of speech from the perspective of the community of speakers in their own ideological context, rather than as an observer "looking in" and applying broad categories. (Bucholtz and Hall, 2004; Bucholtz and Hall, 2005).

Fractal Recursivity The semiotic process by which a salient opposition is projected onto other levels or types of relationships (Irvine, Gal, and Kroskrity, 2009). For example, the projection of urban lifestyles as "rushed" and rural lifestyles as "slow" being projected onto speech, and speech rate differences between populations (whether they actually exist or not) are taken to be a reflection of that lifestyle difference (Evans, 2013).

Gender The fuzzy ideological category that refers to behavioural, cultural, or social expectations of individuals that are assumed to follow from their biological/reproductive status.

> "Gender... is not something residing inside the human, but a relational concept, just like 'big' cannot be 'big' unless there is something other than 'big', that makes it so." (Gherardi, 1995)

While I recognize there are nuances in source and appeal to authority between the definitions of sex and gender, I collapse them here. Throughout this thesis, unless explicitly outlined (or attributed to another author), please read sex, gender, and sex/gender as equivalent terms. Likewise, words that are sometimes used to index one or the other meaning exclusively (such as *woman* and *female*) are used interchangeably in this thesis.

Iconisation The semiotic process by which a linguistic feature becomes associated with a social category (Irvine, Gal, and Kroskrity, 2009). For example, the way high-frequency spectral peaks in /s/ production are linked to the speech of gay men, whereby listeners rank instances of that kind of /s/ as "sounding gay" (Munson and Babel, 2007).

Indexical Order A hierarchical ranking of the relationship between (linguistic) forms and social categories. First-order indexicality refers to relationships that are recognized by outsiders as associated with macro-social categories, second-order

indexicality to relationships recognized by insiders as reflecting ideological stances, and third-order indexicality refers to relationships that are connotative within the moment/context of the speech act (Silverstein, 2003).

Indexicality The link between speech properties and identity. "A property of speech through which cultural contexts such as social identities (e.g. gender) and social activities (e.g. gossip session) are constituted by particular stances and acts" Ochs (1992, p. 335).

Individual Vernacular An idiolect. The way a given person speaks, assumed to be unconscious, and separate from the standard dialect of the area/group (Labov, 1968; Labov, 1972).

Intersectionality (Also Multiplicity). The interaction of different aspects of identity (such as race, gender, and disability), and the ways those interactions in perception and performance create unique stereotypes, interactions, and ways of being or being perceived, that are more than the additive sum of parts. Crenshaw (2017).

In the Wild Refers to recordings made by participants in natural, every day settings, without the researcher present. Full recording context for ItW recordings can be found in Appendix C

Incongruence A mismatch between a combination of variables associated with some identity and the actual experience of those variables, arising from the identity itself being unrepresented in the association. See 3.2 for an expanded definition of incongruence with specific reference to gender.

ItW In the Wild

LaBB-CAT Fromont and Hay (2012)'s Language, Brain & Behaviour Corpus Analysis Tool

LGBT+ An acronym describing various aspects of the rainbow community. it stands for **L**esbian, **G**ay, **B**isexual, **T**rans, and Others (**+**). The long-form version of the acronym most commonly seen is **LGBT2QFIA+**, which additionally names **T**wo-Spirit, **Q**ueer/Questioning, **F**a'afafine, **I**ntersex, and **A**sexual/Agender identities. Even within the long form, there are many different versions of the acronym, and disagreement between the specificity/scope of each letter. (For example, the **A** is sometimes taken to stand for *asexual*, *agender*, *ally*, or any combination of those terms.) For the purposes of this thesis, I use **LGBT+** and take it to mean all gender/sexual/attraction minorities that consider themselves part of the rainbow community, and exclude cishet allies.

LGBT+ Lesbian, Gay, Bi, Trans, and Others

MAD Median Absolute Deviation

Multiplicity (Also Intersectionality). The interaction of different aspects of identity (such as race, gender, and disability), and the ways those interactions in perception and performance create unique stereotypes, interactions, and ways of being or being perceived, that are more than the additive sum of parts. Crenshaw (2017), Cameron and Kulick (2003), and Eckert (2008).

Nonbinary An umbrella term encompassing genders outside of the individual binary categories of "man" and "woman." Can also be spelt *Non-binary* or *non binary*. I use the non-hyphenated single-word form here to reduce the morphological emphasis on *non*.

NZE New Zealand English

Orientation / Sexuality The types of bodies and people one is attracted to, both romantically and sexually, often in relation to one's own gender. Note that terms often conflate attraction and gender. For example, the terms *lesbian*, *bisexual*, and *pansexual* differ in scope in the following ways: *Lesbian* is coded as a woman who is attracted to other women, *bisexual* refers to attraction both to genders like one's own (whatever that may be), as well as genders not like one's own, while the term *pansexual* references a potential attraction to all genders without referece to one's own gender.

Performativity The repetition and acting out of the conventions associated with dominant culture identites (particularly gender). The process by which reality and identity is constructed through the iterated acts of doing (performing) that identity. (Butler, 1990).

Personae The combination of speech features and other semiotic cues to create a consistent, contextually grounded identity (Podesva, 2007; Eckert, 2008).

Prestige A concept in linguistic variation whereby different linguistic forms may be evaluated positively (or negatively) by listeners (Trudgill, 1972). Prestige may be overt or covert.

RAINBO Stylized RAINBO throughout, it refers to the Recorded Audiovisual Interviews with Nonbinary and Binary Orators corpus. The RAINBO corpus, as of 2021, contains the recordings and selfies from 16 speakers in Christchurch, New Zealand. It is the dataset used throughout this thesis.

REAPER Talkin (2015)'s **R**obust **E**poch **A**nd **P**itch **E**stimato**R**.

REML Restricted Maximum Likelihood

SES Socioeconomic Status

Sex Fuzzy ideological category often used to refer to broad biological, physiological, chromosomal, or reproductive differences between individuals and the behavioral/social expectations that are assumed to follow from those differences. While I recognize there are nuances in source and appeal to authority between the terms sex and gender, I collapse them here. Throughout this thesis, unless explicitly outlined (or attributed to another author), please read sex, gender, and sex/gender as equivalent terms. Likewise, words that are sometimes used to index one or the other meaning exclusively (such as *woman* and *female*) are used interchangeably in this thesis.

Social Networks The relationships between speakers in a speech community and beyond. The "web of ties" connecting speakers to each other. (Milroy, 1980; Milroy and Llamas, 2013).

Social Stratification The organization of people in a society into different groups/rankings based on relative wealth, power, or other categories. (Labov, 2006; Trudgill, 2004).

Sociolinguistic Interview A method for collecting linguistic data from speakers that utilizes varying levels of formal structure, intended to elicit variation between standard and more local (vernacular) ways of speaking. (Labov, 2006; Tagliamonte, 2006).

Style, Stance Ways of speaking which are associated with social features. Stances are small instances of variation that can be used in combination to perform a speech style. (Labov, 1972; Bucholtz and Hall, 2005).

Transgender (or *trans*) Someone whose gender is different to the one assigned to them at birth. The opposite of cis.

VD Vowel Dispersion

VSA Vowel Space Area

VSC Vowel Space Conformity

VTL Vocal Tract Length

Appendix B

Research Materials

This Appendix Contains the materials used in recruiting and recording participants for inclusion in the corpus of Recorded Audiovisual Interviews with Nonbinary and Binary Orators . The demographic questionnaire participants were initially asked to complete is in Section B.1. The information they were asked to provide about each recording is in Section B.2. The complete interview questions, as well as the Topic/Target/Time tagging for each question is in Section B.3.

B.1 Demographic Questions

Participant Number:

Department: Linguistics
Telephone: +64 3 364 2987 ext 8862
Email: jacq.jones@pg.canterbury.ac.nz

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Linguistics Effects of Gender in Different Conversational Contexts

Demographic Questionnaire

1. **How old are you?** _____
2. **What is your gender / gender identity?** _____
3. **What is your occupation?** _____
4. **Mother's occupation:** _____
5. **Father's occupation:** _____
6. **Have you ever been diagnosed with a hearing impairment?** Yes / No
7. **Are you currently a student at Canterbury?** Yes / No
8. **Are you:** Right-handed / Left-handed
9. **Total years spent living <u>outside</u> New Zealand:** less than 2 2–4 4–6 6 or more

   (List countries, other than New Zealand, that you have lived in for more than a year, and how long)

   Ex: USA, 10 years _____ _____

   _____ _____ _____
10. **Are you fully fluent in any languages other than English?** Yes / No

    (list language, rate fluency from 1-3 where 1 = rudimentary and 3 = fluent, and list age learned)

    Ex: French, 3, 6 years _____ _____

    _____ _____ _____

B.2 Recording Slips

Participant Number: _____

Recording #1:

When was this recorded?

Morning Afternoon Evening

Date: _____

How many people were in this conversation (including yourself)?

1 2 3 More than 3

What is your relationship to the other people in the conversation?
(e.g. friends, family, coworkers, boss)

_____

Where was this recorded?
(e.g. home, work, a park)

_____

What is/are the general topic(s) of conversation?
(e.g. work, the weather, future plans)

_____

Please provide the filename (or brief description) of the selfie that accompanies this recording:

_____

_____

Recording #2:

When was this recorded?

Morning Afternoon Evening

Date: _____

How many people were in this conversation (including yourself)?

1 2 3 More than 3

What is your relationship to the other people in the conversation?

_____

_____

Where was this recorded?

_____

_____

What is/are the general topic(s) of conversation?

_____

_____

Please provide the filename (or brief description) of the selfie that accompanies this recording:

_____

_____

Recording #3:

When was this recorded?

Morning Afternoon Evening

Date: _____

How many people were in this conversation (including yourself)?

1 2 3 More than 3

What is your relationship to the other people in the conversation?

_____

_____

Where was this recorded?

_____

_____

What is/are the general topic(s) of conversation?

_____

_____

Please provide the filename (or brief description) of the selfie that accompanies this recording:

_____

_____

B.3 Interview Questions and Topic Tagging

Question	Question Code	Broad Tag	Target Tag	Time Tag
Per Recording				
What is this conversation about?	RWhat			Near Past
Where did this conversation take place?	RWhere			Near Past
Who was this conversation with? Are you close to them?	RWho	*variable*	*variable*	Near Past
(Nonbinary/LGBT+ Participants only)				
Does the person you're speaking to know you're <gender>?	Rout	Gender	*variable*	Near Past
How did you feel during this conversation?	RFeel	*variable*	Self	Near Past
Do you remember thinking about your physical appearance at all?	RSee	Physical	Self	Near Past
Do you remember making any decisions about the way you spoke?	RSpeak	Language	Self	Near Past
Tell me about this outfit.	Selfie	Physical	Self	Near Past
Is there's a story behind this clothing?	SMem	Physical	Self	Diachronic
Ask about anything that stands out, such as shoes, makeup, jewelry, etc.	Sother	Physical	Self	
Questions About the Participant's Gender and Awareness of Nonbinary Genders				
What is your gender?	SRGender	Gender	Self	Present
(Binary Participants only)				
How do you feel about being <gender>?	Cgender	Gender	Self	Present
Have you heard of people who aren't male/female? What do you think about that?	SRawareness	Gender	*variable*	Present
(Nonbinary Participants only)				
Could you tell me about your "journey", when you realised you might not be binary, the first time you put a name to your gender, that sort of thing?	SRJourney	Gender	Self *variable*	Far Past
Has your gender changed over time?	SRDiachronic	Gender	Self	Diachronic

Question	Question Code	Broad Tag	Target Tag	Time Tag
Do you think you can "tell" when someone's nonbinary? How?	SRradar	Gender Physical	Theoretical	Present
Questions about Body, Body Image, and Physical Manifestations of Gender				
What's your favourite part of your body? Why?	SRbodypos	Physical	Self	Present
Do you do anything to show that part of your body off?	SRshow	Physical	Self	Present
What's your least favourite part of your body? Why?	SRbodyneg	Physical	Self	Present-
Do you do anything to hide that part of your body?	SRhide	Physical	Self	Present
If you could change one thing about your body, what would it be?	SRbodywish	Physical	Self	Fut/Spec
How important is physical appearance to gender presentation?	SRimport	Physical Gender	Theoretical	Present
How does physical appearance alter people's perception of gender?	SRassumption	Physical Gender	Theoretical	Present
Do you "use" your physical appearance for presenting your gender?	SRuse	Physical Gender	Self	Present
(Nonbinary Participants only)				
Have you ever tried to change your body to match your gender?	SRbodychange	Physical Gender	Self	Near Past
Have you sought intervention/aid via medical channels for presenting or affirming your gender?	SRmed	Physical Gender	Self	Near Past
Have you sought intervention/aid via nonmedical channels for presenting or affirming your gender?	SRnonmed	Physical Gender	Self	Near Past
Questions about Family				
How do you define family?	Fdefine	Family	Theoretical	Present

Question	Question Code	Broad Tag	Target Tag	Time Tag
If you could tell your parents' generation one thing about being <gender> in <year>, what would it be?	Fmessage	Family	Theoretical Close Other	Fut/Spec
(Older Participants only) If you could tell your kids' / younger generation one thing about being <gender> in <year>, what would it be?	Fmessage	Family Gender	Theoretical Close Other	Fut/Spec
(Nonbinary Participants only) Who is your family?	Fwho	Family	Close Other	Present
Are you "out" to your family?	Fout	Family	Self	Near Past
Has your relationship with your family changed over time?	Fdiachronic	Family	Self	Diachronic
Questions about Language				
How do <gender> speak English?	Lgrouptrend	Language Gender	Theoretical	Present
Do you think gender plays a role in how people sound? How? Why?	Lgender	Language Gender	Theoretical	Present
Do you speak the same way as your friends?	Lfriends	Language	Close Other	Present
Are there words you use that other people don't use?	Lwords	Language	Self	Present
Have you ever tried to change the way you talk?	Lchange	Language	Self	Diachronic
Has anyone ever given you a hard time about the way you talk?	Lhard	Language	*variable*	Near Past
Has anyone every said to you, "You sound different"?	Ldifferent	Language	*variable*	Near Past
Do you think how you sound changes others' perception of you?	Lperception	Language Gender	Theoretical	Present
Do you try to change how you sound in different environments?	Lstyle	Language	Self	Diachronic

Question	Question Code	Broad Tag	Target Tag	Time Tag
How important is language to gender presentation?	Limport	Language Gender	Theoretical	Present
How does language alter people's perception of gender?	Lassumption	Language Gender	Theoretical	Present
Do you "use" your language or the way that you talk for presenting your gender? How?	Luse	Language Gender	Self	Present
How would you describe how you sounded to a person who had never heard your voice?	Lstranger	Language Gender	Theoretical	Fut/Spec
Meta Questions				
What does masculinity mean to you, how would you define it? What's your relationship to masculinity?	Mmasc	Gender *variable*	Theoretical	Present
What does femininity mean to you, how would you define it? What's your relationship to femininity?	Mfemme	Gender *variable*	Theoretical	Present
What were you told about the difference between "boys" and "girls" growing up?	Mchildhood	Gender Physical *variable*	*variable*	Far Past
What do you think I'm measuring?	Meta	Gender Meta	Theoretical	Present
(Nonbinary Participants only) What is gender?	Mgender	Gender	Self	Present
Is there anything you'd like to tell me about the ways other aspects of your identity interact with your gender? (For example, sexuality, race/ethnicity, ability, religion, etc).	Mintersect	Gender	Self	Present

Appendix C

In the Wild Recording Information

This section contains a brief description of each participant's ItW recordings. Information here has been generalised (for example, "café" instead of "Starbucks") to maintain anonymity. "Tlk" is short for "Talkers", and refers to the number of interlocutors (including the participant) present in each recording.

Alex

Rec.	Tlk.	Time	Relation	Location	Topics
1	3+	Afternoon	Colleagues	Work	Work
2	2	Afternoon	Friend	Home	Hobbies
3	2	Morning	Friend	Public Space	Kids, Hobbies
4	3+	Afternoon	Children	Home	Math
5	2	Afternoon	Employee	Garden	Work
6	2	Evening	Friend	Public Space	Food, Event

TABLE C.1: ItW Recording Details: Alex

Avery

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Family	Vehicle	Event, Hobbies
2	2	Morning	Friend	Home	Friendship, Study
3	3	Evening	Friend	Home	Catching up, Hobbies

TABLE C.2: ItW Recording Details: Avery

Devon

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Afternoon	Friend	Home	Hobbies, History
2	2	Afternoon	Colleague	Public Space	Study, Hobbies
3	2	Afternoon	Colleague	Public Space	Study, Catching up
4	2	Evening	Friend	Home	Sport, Flat

TABLE C.3: ItW Recording Details: Devon

Istus

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Friend	Home	Hobbies
2	2	Afternoon	Friend	Uni	Hobbies
3	2	Afternoon	Colleague	Uni	Study

TABLE C.4: ItW Recording Details: Istus

Jesse

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Afternoon	Family	Café	Family
2	2	Morning	Flatmate	Home	Flat
3	2	Morning	Flatmate	Home	Sport
4	4	Afternoon	Flatmates	Home	Catching up, Hobbies
5	3	Afternoon	Flatmates	Home	Hobbies, Flat
6	2	Evening	Family	Home	Catching up, Family

TABLE C.5: ItW Recording Details: Jesse

Jordon

Rec.	Tlk.	Time	Relation	Location	Topics
1	3	Afternoon	Friends	Home	Event, Catching up
2	3+	Evening	Colleagues	Home	Media
3	3+	Evening	Colleagues	Home	Catching up

TABLE C.6: ItW Recording Details: Jordon

Kelly

Rec.	Tlk.	Time	Relation	Location	Topics
1	3+	Evening	Family	Home	Catching up, Kids
2	2	Morning	Friend	Café	Catching up, Computers
3	2	Afternoon	Friend	Café	Kids, School
4	2	Morning	Family	Café	Future plans, Work
5	2	Afternoon	Friend	Public Space	Art, Catching up, Work

TABLE C.7: ItW Recording Details: Kelly

Kit

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Family	Vehicle	Study, Hobbies
2	4+	Evening	Friend	Home	Memes
3	3	Evening	Family	Public Space	Study, Hobbies
4	2	Evening	Family	Home	Media, Philosophy
5	4+	Evening	Guests	Home	Food, Language
6	2	Morning	Friend	Café	Study, Hobbies

TABLE C.8: ItW Recording Details: Kit

Mello

Rec.	Tlk.	Time	Relation	Location	Topics
1	1	Afternoon	Flatmate	Home	Catching up
2	2	Morning	Family	Vehicle	Family, Work
3	2	Afternoon	Friend	Public Space	Sport

TABLE C.9: ItW Recording Details: Mello

Noel

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Flatmate	Home	History
2	2	Evening	Friend	Home	Dating
3	2	Evening	Family	Home	Flat

TABLE C.10: ItW Recording Details: Noel

Ollie

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Afternoon	Family	Vehicle	Food, Study
2	2	Evening	Family	Home	Pets
3	2	Evening	Family	Home	Catching up
4	3+	Evening	Family	Home	Food, Pets
5	2	Morning	Friend	Vehicle	Study
6	3+	Afternoon	Friends	Vehicle	Study, Friends

TABLE C.11: ItW Recording Details: Ollie

Quinn

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Afternoon	Friends	Home	Chores
2	3+	Afternoon	Colleagues	Uni	Club Activities
3	2	Afternoon	Friends	Uni	Hobbies, Event

TABLE C.12: ItW Recording Details: Quinn

Reese

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Flatmate	Home	Work, Travel, Hobbies
2	3	Evening	Clients	Work	Weather, Hobbies
3	2	Afternoon	Family	Home	Study

TABLE C.13: ItW Recording Details: Reese

Remi

Rec.	Tlk.	Time	Relation	Location	Topics
1	3	Evening	Family	Home	Catching up
2	4+	Afternoon	Friends	Uni	Club Activities
3	2	Evening	Colleagues	Home	Catching up

TABLE C.14: ItW Recording Details: Remi

Riley

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Friend	Uni	Food, Study
2	2	Afternoon	Colleague	Uni	Food, Study
3	3	Morning	Friend	Home	Event
4	2	Afternoon	Stranger	Home	Phone Call
5	5	Afternoon	Family	Home	Catching up, Family

TABLE C.15: ItW Recording Details: Riley

Taylor

Rec.	Tlk.	Time	Relation	Location	Topics
1	2	Evening	Friend	Home	Movies, Hobbies
2	5	Evening	Friends	Home	Uni, Argument
3	2	Afternoon	Family	Home	Family, Catching up
4	3	Morning	Friends	Uni	Study
5	4	Evening	Family	Home	Catching up, Study
6	2	Evening	Friend	Home	Catching up, Food

TABLE C.16: ItW Recording Details: Taylor

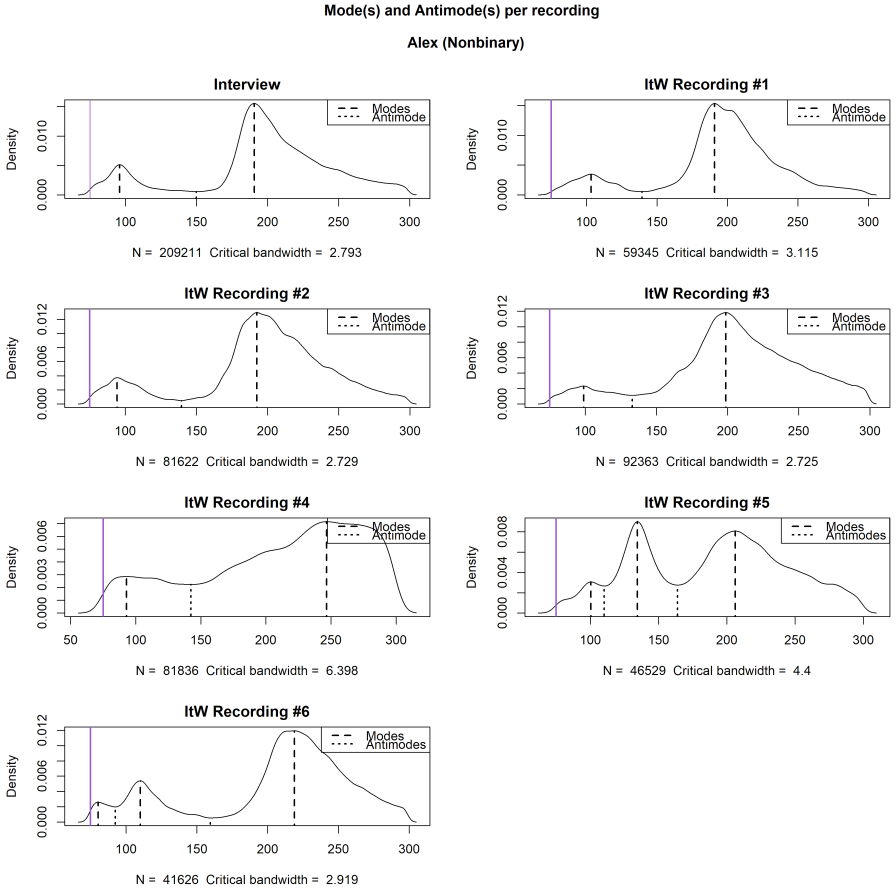
Appendix D

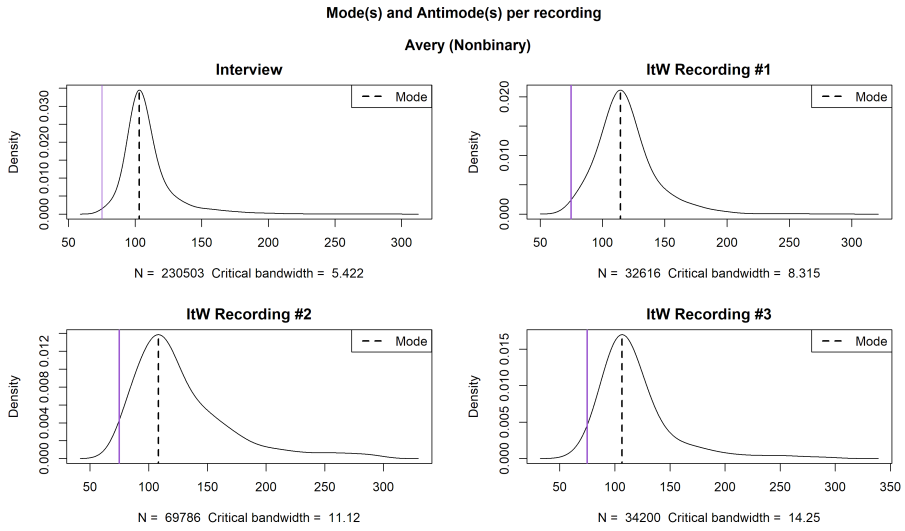
Pitch Bins (Modes and Antimodes)

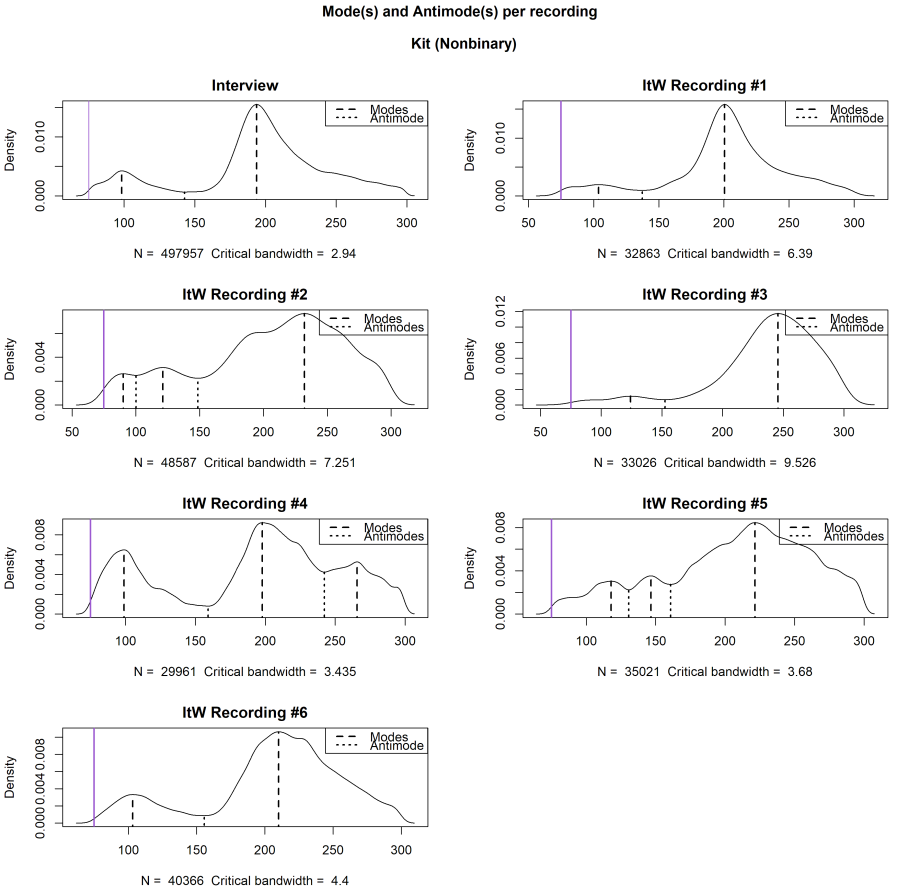
The following pages contain "Pitch Bin" visualisation of each recording for each participant, including modes and antimodes as estimated by the `locmodes` function of the multimode package in R (Ameijeiras-Alonso, Crujeiras, and Rodríguez-Casal, 2021). The x-axis of each graph is the measurement of F0 (in Hz), while the y-axis shows the density. The purple line on the leftward side is the 75Hz mark, below which values were removed as likely being produced with creaky phonation. The number of modes and antimodes per recording was determined via a combination of visual analysis and the `modetest` function of the same package. the `modetest` function was first used to determine the likelihood that the distribution was unimodal ($p \geq 0.5$). If the distribution was found to contain more than one mode, visual analysis was used to determine how many modes were likely, and subsequently plotted, in each recording.

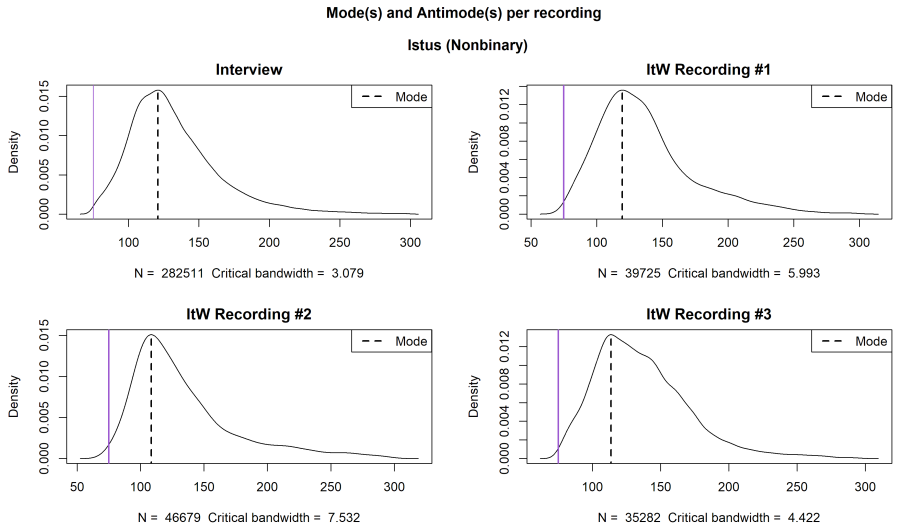
Following the visualisations is Table D.1, containing the summary statistics of each recording for each speaker. Unimodal distributions have Mode 1 listed as the most frequent value as determined by the Modeest R package (Poncet, 2019). In cases where multiple modes have been identified, I have also listed the second (and sometimes third) mode as identified by the `locmodes` function. Modes differ slightly from the ones visualised, a few Hz at most, firstly because the `locmodes` function determines multiple modes based on a complex density calculation, while the `mfv` function simply takes the most frequent value, and secondly because the visualisations were calculated with creak removed, subtly altering the density estimates. In cases where 2.5 M.A.D. resulted in values beyond REAPER's capability to measure, I changed these to reflect REAPER's floor and ceiling (40.201Hz and 296.296Hz, respectively).

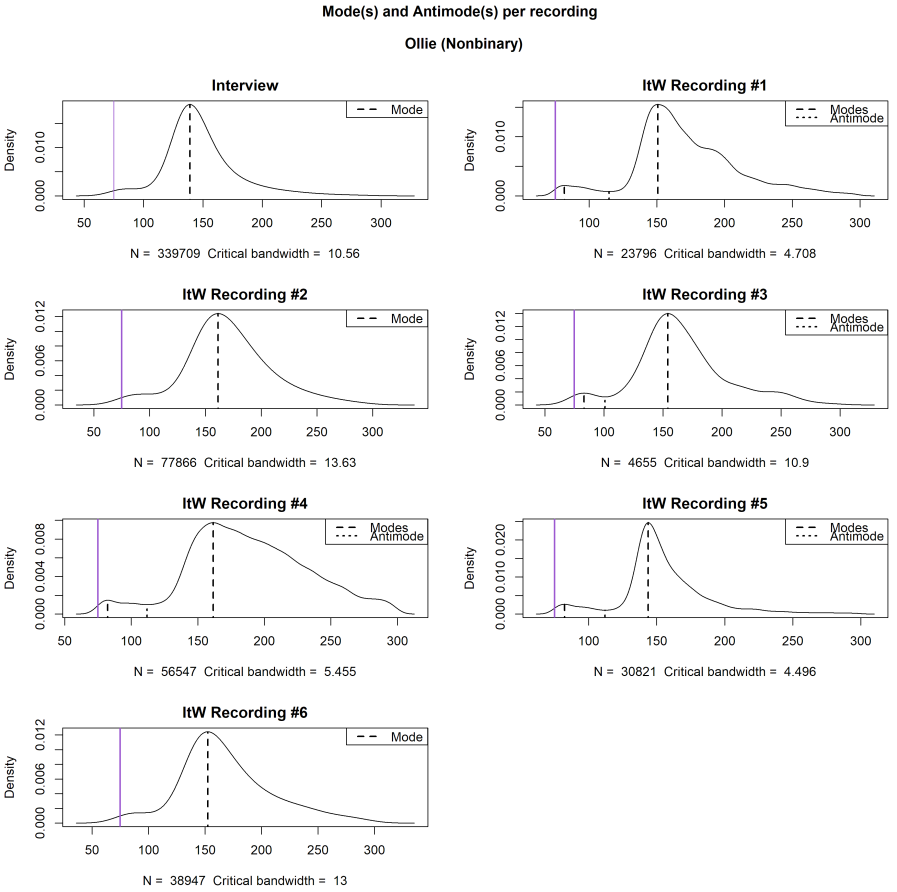
D.1 Nonbinary Speakers

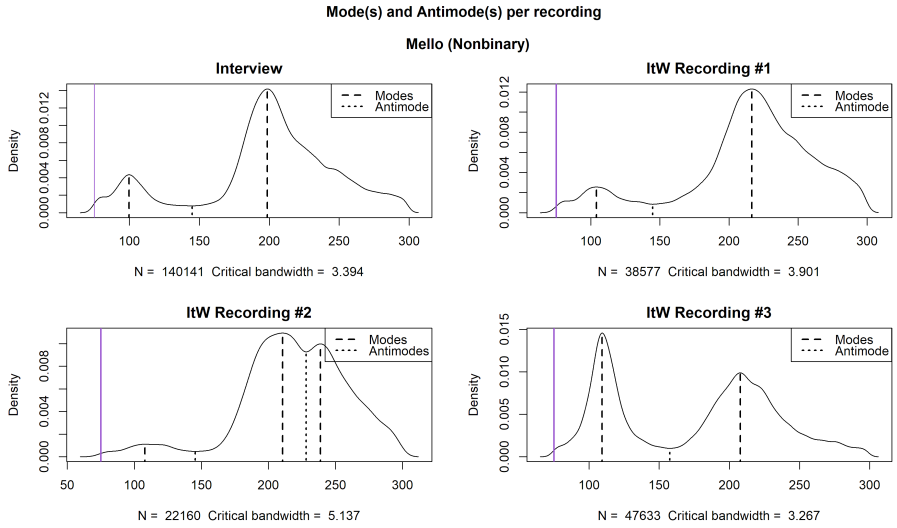



Mode(s) and Antimode(s) per recording
Avery (Nonbinary)
Interview
Mode
Density
N = 230503 Critical bandwidth = 5.422
ItW Recording #1
N = 32616 Critical bandwidth = 8.315
ItW Recording #2
N = 69786 Critical bandwidth = 11.12
ItW Recording #3
N = 34200 Critical bandwidth = 14.25

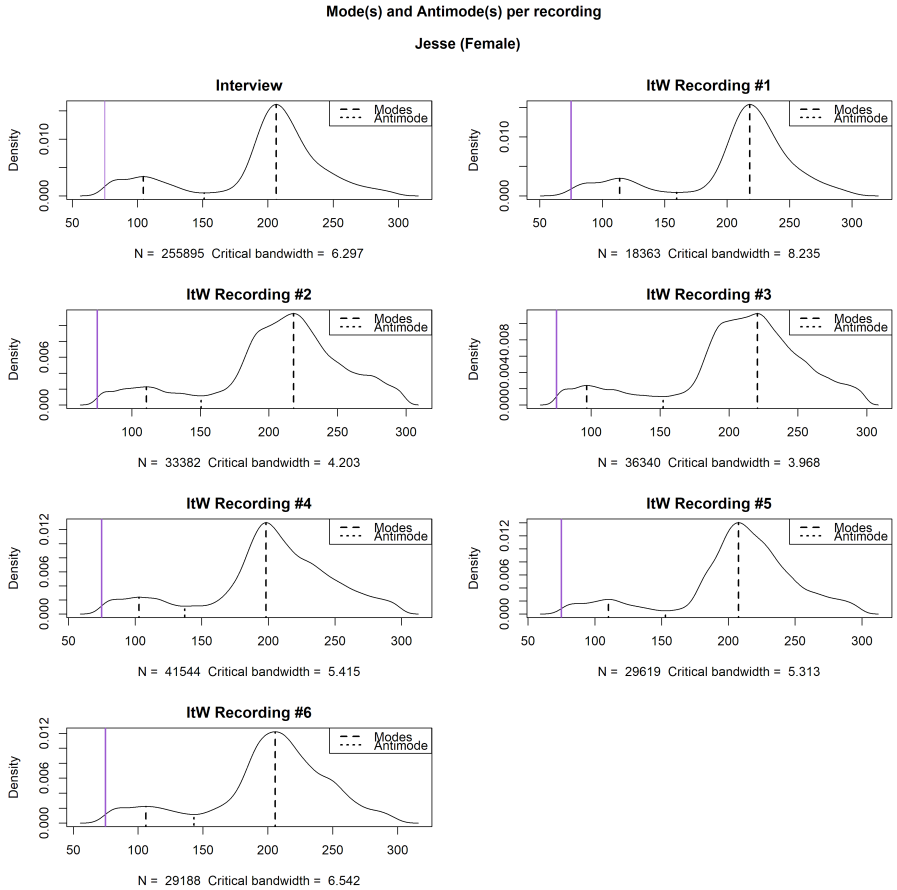

Mode(s) and Antimode(s) per recording
Kit (Nonbinary)
Interview
Modes
Antimode
Density
N = 497957 Critical bandwidth = 2.94
ItW Recording #1
N = 32863 Critical bandwidth = 6.39
ItW Recording #2
Antimodes
N = 48587 Critical bandwidth = 7.251
ItW Recording #3
N = 33026 Critical bandwidth = 9.526
ItW Recording #4
N = 29961 Critical bandwidth = 3.435
ItW Recording #5
N = 35021 Critical bandwidth = 3.68
ItW Recording #6
N = 40366 Critical bandwidth = 4.4

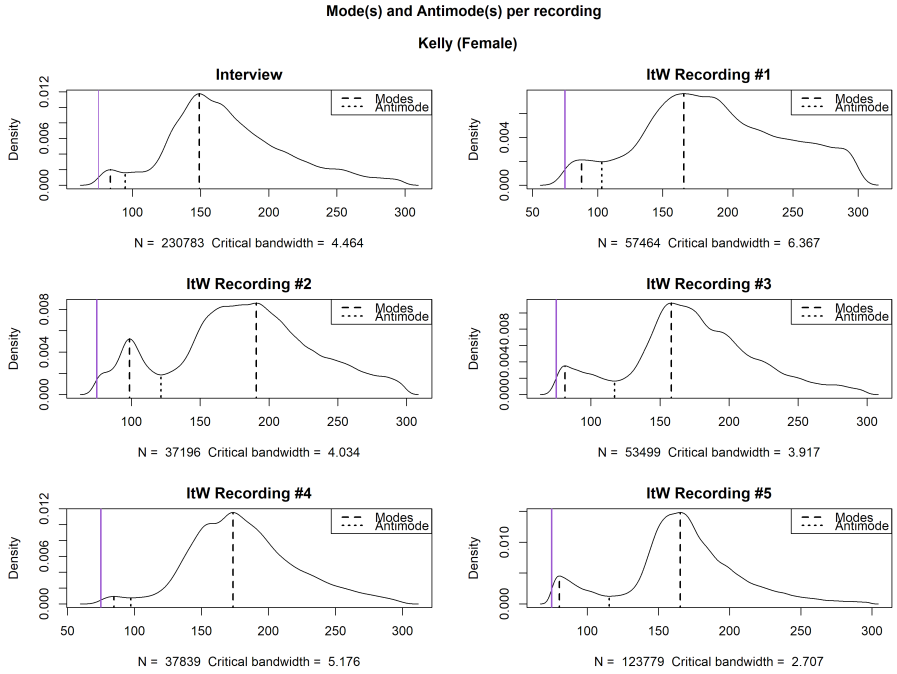

Mode(s) and Antimode(s) per recording
Istus (Nonbinary)
Interview
Mode
Density
N = 282511 Critical bandwidth = 3.079
ItW Recording #1
N = 39725 Critical bandwidth = 5.993
ItW Recording #2
N = 46679 Critical bandwidth = 7.532
ItW Recording #3
N = 35282 Critical bandwidth = 4.422

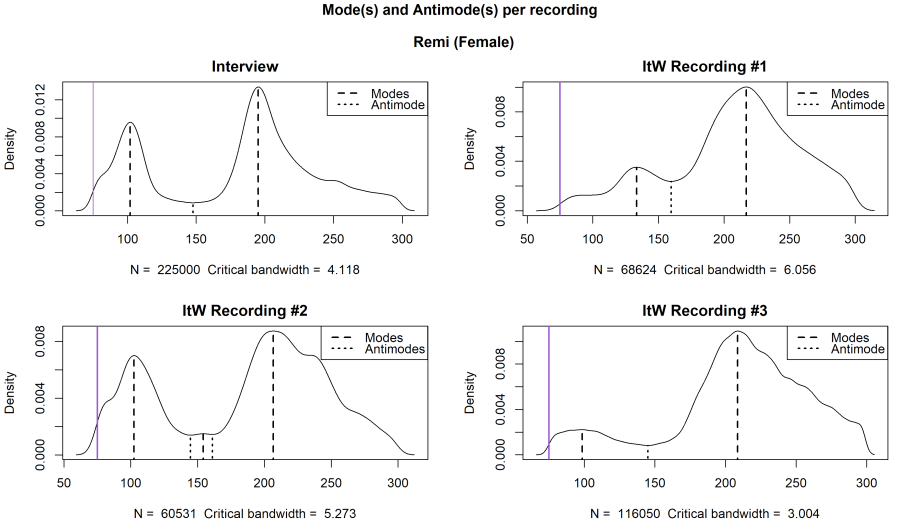

Mode(s) and Antimode(s) per recording
Ollie (Nonbinary)
Interview
Mode
Density
N = 339709 Critical bandwidth = 10.56
ItW Recording #1
Modes
Antimode
N = 23796 Critical bandwidth = 4.708
ItW Recording #2
N = 77866 Critical bandwidth = 13.63
ItW Recording #3
N = 4655 Critical bandwidth = 10.9
ItW Recording #4
N = 56547 Critical bandwidth = 5.455
ItW Recording #5
N = 30821 Critical bandwidth = 4.496
ItW Recording #6
N = 38947 Critical bandwidth = 13

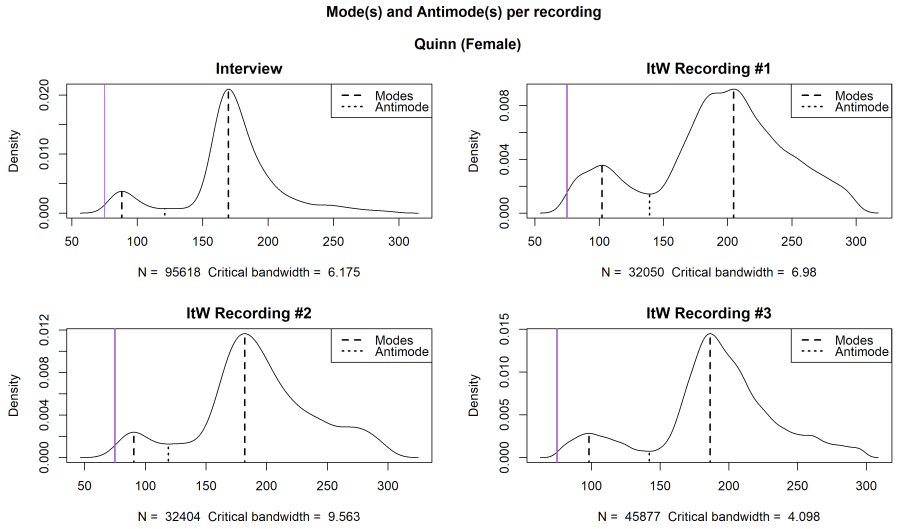


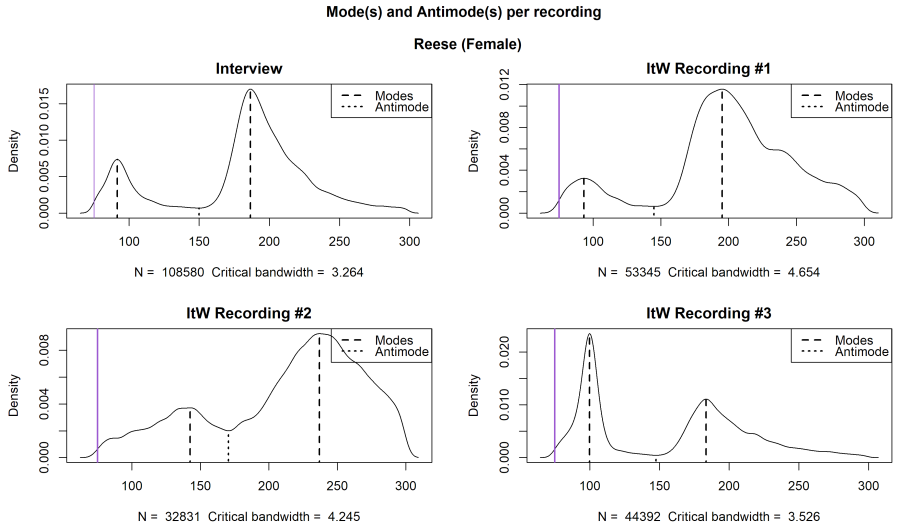
D.2 Female Speakers



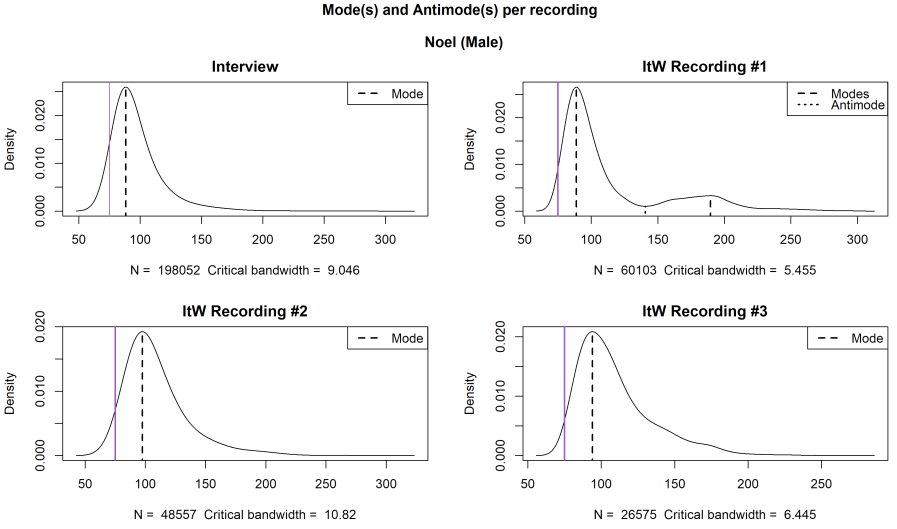

Mode(s) and Antimode(s) per recording
Kelly (Female)
Interview
Modes
Antimode
Density
N = 230783 Critical bandwidth = 4.464
ItW Recording #1
N = 57464 Critical bandwidth = 6.367
ItW Recording #2
N = 37196 Critical bandwidth = 4.034
ItW Recording #3
N = 53499 Critical bandwidth = 3.917
ItW Recording #4
N = 37839 Critical bandwidth = 5.176
ItW Recording #5
N = 123779 Critical bandwidth = 2.707

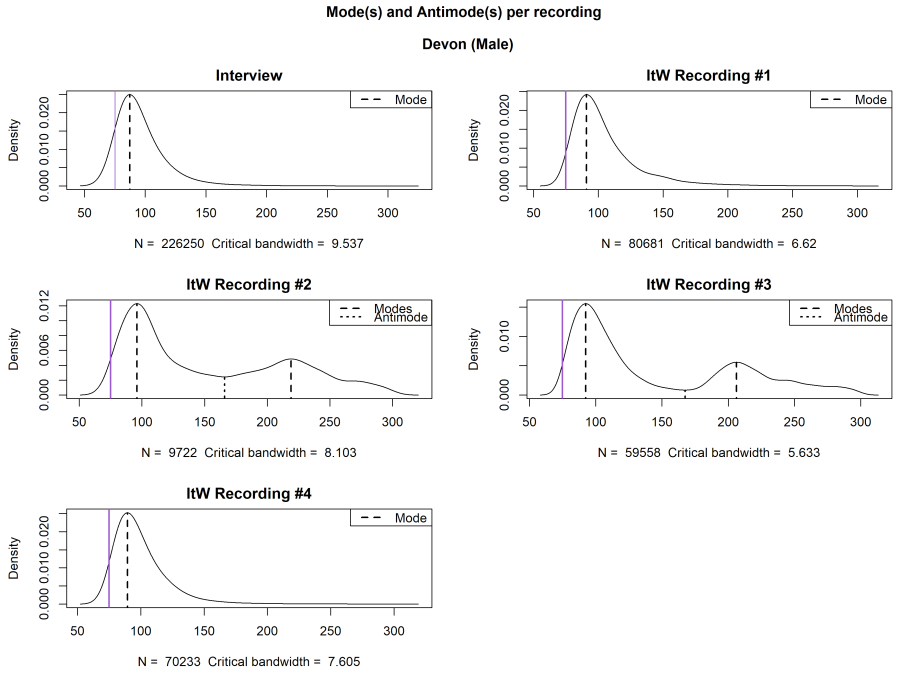

Mode(s) and Antimode(s) per recording
Remi (Female)
Interview
Modes
Antimode
Density
N = 225000 Critical bandwidth = 4.118
ItW Recording #1
N = 68624 Critical bandwidth = 6.056
ItW Recording #2
Antimodes
N = 60531 Critical bandwidth = 5.273
ItW Recording #3
N = 116050 Critical bandwidth = 3.004

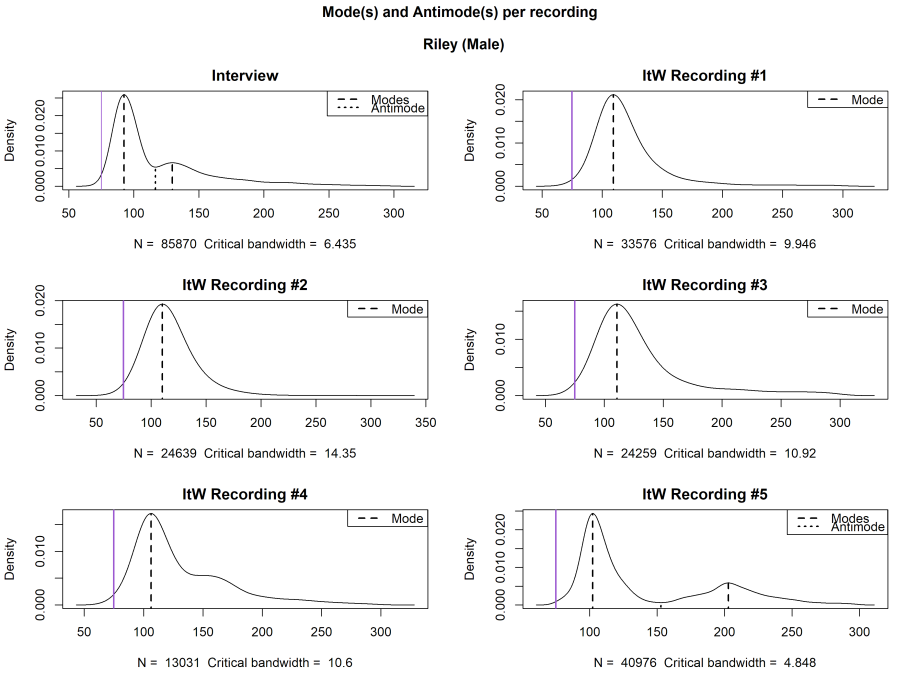


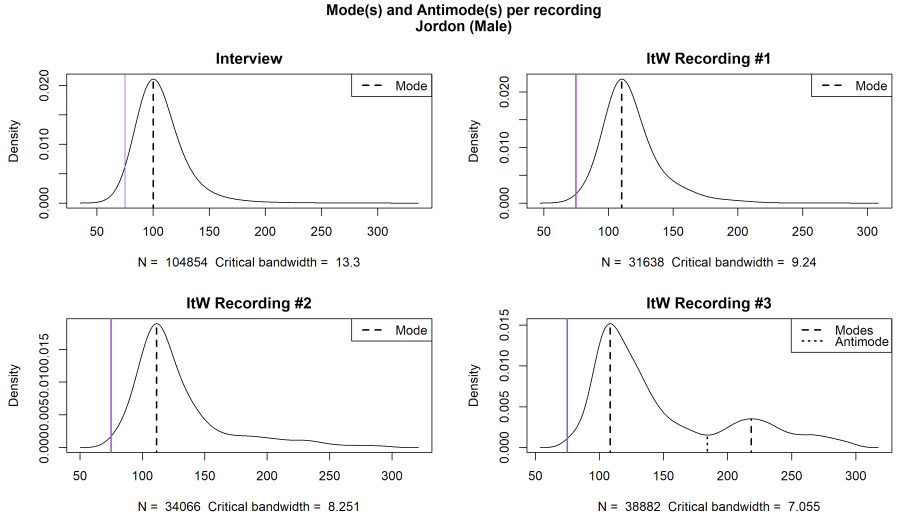


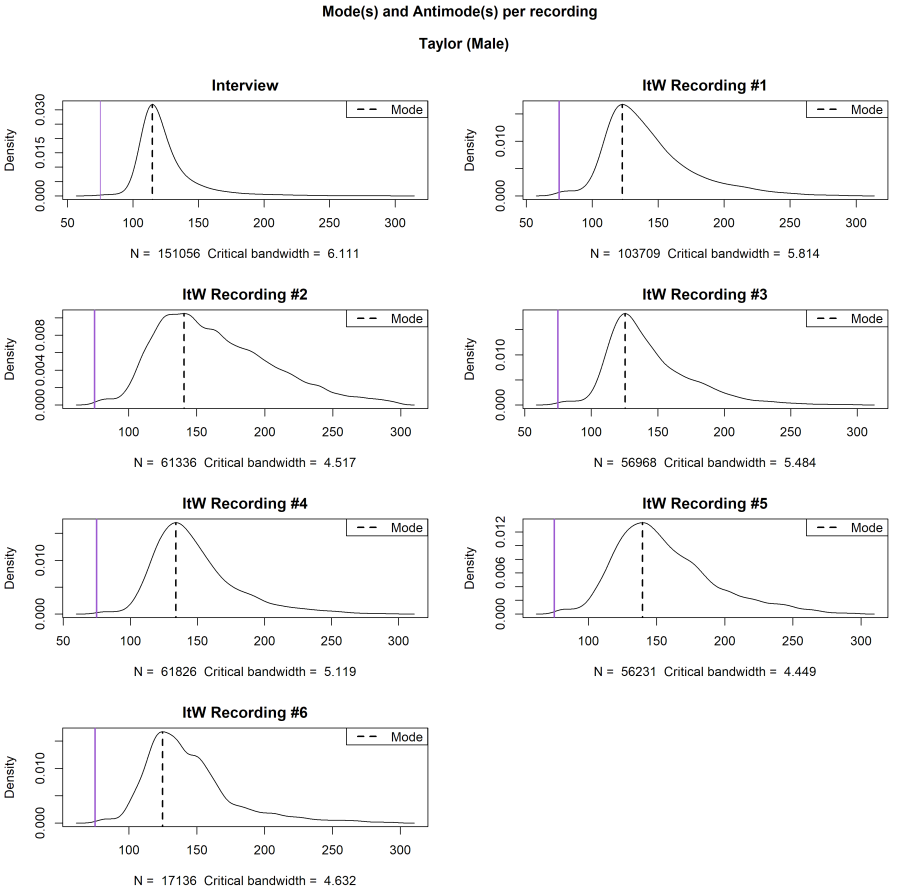
D.3 Male Speakers




Mode(s) and Antimode(s) per recording
Devon (Male)
Interview
Mode
Density
N = 226250 Critical bandwidth = 9.537
ItW Recording #1
Mode
N = 80681 Critical bandwidth = 6.62
ItW Recording #2
Modes
Antimode
N = 9722 Critical bandwidth = 8.103
ItW Recording #3
Modes
Antimode
N = 59558 Critical bandwidth = 5.633
ItW Recording #4
Mode
N = 70233 Critical bandwidth = 7.605


Mode(s) and Antimode(s) per recording
Riley (Male)
Interview
Modes
Antimode
Density
N = 85870 Critical bandwidth = 6.435
ItW Recording #1
Mode
N = 33576 Critical bandwidth = 9.946
ItW Recording #2
Mode
N = 24639 Critical bandwidth = 14.35
ItW Recording #3
Mode
N = 24259 Critical bandwidth = 10.92
ItW Recording #4
Mode
N = 13031 Critical bandwidth = 10.6
ItW Recording #5
Modes
Antimode
N = 40976 Critical bandwidth = 4.848


Mode(s) and Antimode(s) per recording
Jordon (Male)
Interview
Mode
Density
N = 104854 Critical bandwidth = 13.3
ItW Recording #1
N = 31638 Critical bandwidth = 9.24
ItW Recording #2
N = 34066 Critical bandwidth = 8.251
ItW Recording #3
Modes
Antimode
N = 38882 Critical bandwidth = 7.055


Mode(s) and Antimode(s) per recording
Taylor (Male)
Interview
Mode
Density
N = 151056 Critical bandwidth = 6.111
ItW Recording #1
N = 103709 Critical bandwidth = 5.814
ItW Recording #2
N = 61336 Critical bandwidth = 4.517
ItW Recording #3
N = 56968 Critical bandwidth = 5.484
ItW Recording #4
N = 61826 Critical bandwidth = 5.119
ItW Recording #5
N = 56231 Critical bandwidth = 4.449
ItW Recording #6
N = 17136 Critical bandwidth = 4.632

D.4 Mode Table

Name *Gender*	Recording	Observations	Mean	M.A.D.	Min	Max	Range	Multimodal?	Mode 1	Mode 2	Mode 3	Range Cross-Context
Alex	Interview	226,011	186.10	36.61	94.58	277.62	183.04	Yes	192.77	82.29		
Nonbinary	1	62,149	188.29	29.29	115.08	261.51	146.43	Yes	202.53	103.80		
	2	89,555	184.62	37.79	90.14	279.09	188.94	Yes	200.00	99.26		
	3	95,840	198.67	42.43	92.60	296.30	203.70	Yes	200.00	104.39		
	4	95,277	188.64	87.29	40.20	296.30	256.10	Yes	285.71	94.52		
	5	50,374	178.27	72.80	40.20	296.30	256.10	Yes	210.53	135.93	244.73	
	6	55,646	166.73	82.33	40.20	296.30	256.10	Yes	228.57	111.47	80.26	
												92.94
Avery	Interview	257,083	103.94	11.85	74.31	133.56	59.24	No	103.90			
Nonbinary	1	50,168	95.04	38.88	40.20	192.25	152.05	No	117.65			
	2	82,712	117.45	35.37	40.20	205.88	165.68	No	112.68			
	3	41,654	107.62	22.29	51.88	163.36	111.47	No	107.38			
												13.75
Devon	Interview	259,313	91.91	14.06	56.77	127.05	70.28	No	85.56			
Male	1	89,302	99.91	17.78	55.46	144.35	88.90	No	89.89			
	2	11,530	137.60	61.19	40.20	290.57	250.37	Yes	93.49	222.22		
	3	71,451	129.96	43.00	40.20	237.47	197.27	Yes	93.26	205.13	296.30	
	4	84,508	92.81	17.07	50.13	135.49	85.36	No	89.89			
												7.93
Istus	Interview	305,360	126.66	28.85	54.55	198.77	144.23	No	125.00			
Nonbinary	1	42,110	132.09	31.38	53.65	210.53	156.89	No	137.93			
	2	49,868	128.86	30.60	52.36	205.35	152.99	No	126.98			
	3	36,852	133.24	32.49	52.03	214.46	162.43	No	145.45			
												20.45
Jesse	Interview	277,008	185.84	29.98	110.88	260.80	149.92	Yes	207.79	106.90		
Female	1	19,823	195.95	28.28	125.24	266.66	141.42	Yes	216.22	112.18		
	2	35,466	198.13	39.46	99.47	296.79	197.32	Yes	222.22	115.21		
	3	38,953	196.93	39.46	98.27	295.59	197.32	Yes	225.35	96.48		
	4	45,497	188.54	42.36	82.64	294.44	211.80	Yes	197.53	107.72		
	5	30,756	201.92	31.66	122.76	296.30	173.54	Yes	207.79	106.16		
	6	31,060	192.81	37.59	98.84	296.30	197.46	Yes	216.22	125.71		
												27.82
Jordon	Interview	114,532	102.37	13.11	69.59	135.15	65.56	No	96.39			
Male	1	37,313	107.88	18.60	61.37	154.39	93.02	No	112.68			
	2	43,902	111.53	29.58	40.20	185.47	145.27	No	112.68			
	3	44,031	136.11	34.59	49.64	222.59	172.95	Yes	109.59	216.37		
												16.29
Kelly	Interview	244454	164.26	39.47	65.59	262.92	197.33	Yes	166.67	84.82		
Female	1	66653	171.17	66.24	40.20	296.30	256.10	Yes	190.48	87.12		
	2	42716	165.22	58.44	40.20	296.30	256.10	Yes	192.77	98.76		
	3	64362	153.73	54.46	40.20	289.89	249.69	Yes	177.78	82.59		
	4	38844	178.24	36.89	86.02	270.46	184.44	Yes	170.21	93.43		
	5	138201	155.24	33.78	70.80	239.68	168.88	Yes	170.21	84.65		
												26.1
Kit	Interview	518,221	189.74	32.18	109.29	270.20	160.91	Yes	197.53	99.96		
Nonbinary	1	34,449	194.96	30.71	118.19	271.73	153.55	Yes	202.53	99.78		
	2	55,744	184.46	74.35	40.20	296.30	256.10	Yes	242.42	124.15		
	3	33,894	231.87	35.94	142.01	296.30	154.29	Yes	246.15	125.72		
	4	34,504	174.69	87.25	40.20	296.30	256.10	Yes	198.44	97.01	266.67	
	5	38,051	194.32	62.58	40.20	296.30	256.10	Yes	225.35	119.50		
	6	41,733	201.13	41.93	96.31	296.30	199.99	Yes	228.57	102.17		
												48.62
Mello	Interview	150,452	189.39	37.59	95.42	283.36	187.94	Yes	200.00	99.00		
Nonbinary	1	40,676	206.29	39.15	108.41	296.30	187.89	Yes	222.22	104.01		
	2	22,858	214.18	35.67	125.01	296.30	171.30	Yes	207.93	242.42	104.01	
	3	51,440	166.05	86.76	40.20	296.30	256.10	Yes	210.53	109.49		
												14.32
Noel	Interview	261,146	87.02	16.86	44.88	129.16	84.28	No	86.49			
Male	1	69,094	106.70	17.92	61.89	151.51	89.62	Yes	89.89	188.51		
	2	53,724	104.10	19.98	54.15	154.04	99.89	No	96.97			
	3	29,879	103.70	21.16	50.80	156.60	105.80	No	101.91			
												15.42
Ollie	Interview	372,156	140.14	19.82	90.60	189.69	99.09	No	132.23			
Nonbinary	1	42,800	117.41	106.71	40.20	296.30	256.10	Yes	156.86	82.29		
	2	83,249	163.37	31.28	85.17	241.57	156.40	No	166.67			
	3	5,001	158.41	28.11	88.13	228.69	140.56	Yes	152.38	83.60		
	4	60,501	179.62	45.59	65.65	293.60	227.96	Yes	202.53	84.02		
	5	43,019	125.05	42.43	40.20	231.13	190.92	Yes	142.86	82.57		
	6	43,678	156.30	33.97	71.38	241.23	169.85	No	152.38			
												70.30
Quinn	Interview	103,323	163.68	20.30	112.92	214.45	101.52	Yes	170.21	90.06		
Female	1	41,067	162.38	81.39	40.20	296.30	256.10	Yes	210.53	100.50		
	2	39,864	166.65	56.61	40.20	296.30	256.10	Yes	188.24	92.17		
	3	48,322	184.40	32.70	102.65	266.15	163.49	Yes	188.24	100.98		
												40.32
Reese	Interview	126,696	163.36	47.90	43.61	283.11	239.50	Yes	186.05	91.54		
Female	1	57,954	187.60	41.25	84.48	290.72	206.24	Yes	195.12	91.50		
	2	34,450	206.13	54.77	69.20	296.30	227.1	Yes	246.15	139.70		
	3	55,873	135.72	84.33	40.20	296.30	256.10	Yes	99.46	186.05		
												146.69
Remi	Interview	251,703	166.34	76.33	40.20	296.30	256.10	Yes	102.04	195.12		
Female	1	70,535	203.68	43.63	94.61	296.30	201.69	Yes	219.18	133.85		
	2	86,300	146.28	115.08	40.20	296.30	256.10	Yes	238.81	102.09		
	3	124,863	199.43	42.56	93.02	296.30	203.28	Yes	213.33	101.07		
												136.77
Riley	Interview	113,329	102.42	35.04	40.20	190.01	149.81	Yes	87.91	129.53		
Male	1	45,386	102.70	26.04	40.20	167.80	127.60	No	105.96			
	2	27,715	109.99	15.14	72.14	147.83	75.69	No	106.67			
	3	35,759	104.52	47.22	40.20	222.57	182.37	No	119.40			
	4	13,763	126.84	23.59	67.85	185.82	117.97	No	102.56			
	5	55,764	117.60	50.45	40.20	243.72	203.52	Yes	100.00	202.01		
												31.49
Taylor	Interview	185,864	111.09	14.60	74.60	147.58	72.98	No	112.68			
Male	1	120,017	131.01	28.84	58.91	203.11	144.19	No	124.03			
	2	64,553	158.51	42.77	51.59	265.43	213.84	No	164.95			
	3	60,239	138.87	26.19	73.39	204.35	130.95	No	126.98			
	4	65,565	142.66	26.78	75.70	209.62	133.92	No	139.13			
	5	60,294	149.24	34.73	62.42	236.06	173.64	No	141.59			
	6	19,631	132.83	27.81	63.31	202.35	139.04	No	135.59			
												52.27

TABLE D.1: Summary Stats with Modes. All Measurements in Hz.

Appendix E

Verbatim Model Outputs

E.1 Chapter 6

GLMER reported in 6.6

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: mean_pitch ~ participant_trigen + (1 | who/graph)
   Data: PraatCombo

REML criterion at convergence: 527176

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.8708 -0.5336 -0.1746 0.3627 9.9023

Random effects:
 Groups    Name        Variance Std.Dev.
 graph:who (Intercept) 325.3 18.04
 who       (Intercept) 1074.0 32.77
 Residual              1684.5  41.04
Number of obs: 51309, groups: graph:who, 84; who, 16

Fixed effects:
                       Estimate Std. Error   df t value Pr(>|t|)
(Intercept)              196.19     13.78 12.96   14.24 2.71e-09 ***
participant_trigenFemale 25.58     20.46 13.03    1.25    0.233
participant_trigenMale   -74.43     20.45 12.99   -3.64    0.003 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Correlation of Fixed Effects:
           (Intr) prtc_F
prtcpnt_trF -0.673
prtcpnt_trM -0.674 0.454
convergence code: 0
Model failed to converge with max|grad| = 0.00328052 (tol = 0.002, component
     1)
```

E.2 Chapter 7

Modelling summary from 7.2.3 using normalised data.

Vowel	Model Results			
	F1	*p*	F2	*p*
START	No	0.451	No	0.779
THOUGHT	No	0.576	No	0.643
TRAP	No	0.721	No	0.149
DRESS	No	0.431	No	0.144
FLEECE	No	0.211	No	0.234
KIT	No	0.803	No	0.603
LOT	No	0.197	No	0.486
GOOSE	No	0.952	No	0.874
FOOT	No	0.342	No	0.453
STRUT	No	0.449	No	0.714

TABLE E.1: Results of multiple linear mixed models, examining the relationship between formants and gender in vowel production, using Lobanov-normalised data.

E.3 Chapter 8

GLMER reported in 8.6

```
Generalized linear mixed model fit by maximum likelihood (Laplace
    Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: fricatedbi ~ relevel(trigen, ref = "nonbinary") + freq_scaled +
    sylb_scaled + (1 | speaker/transcript) + (1 | text)
   Data: relevelvtv

     AIC      BIC   logLik deviance df.resid
  1228.7   1270.7   -606.3   1212.7     1403

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.0725 -0.3574 -0.1044 0.3203 6.5747

Random effects:
 Groups           Name        Variance Std.Dev.
 text             (Intercept) 5.289    2.300
 transcript:speaker (Intercept) 0.366 0.605
 speaker          (Intercept) 2.341    1.530
Number of obs: 1411, groups: text, 417; transcript:speaker, 74; speaker, 16

Fixed effects:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                            -0.6480    0.6817 -0.951 0.341796
relevel(trigen, ref = "nonbinary")female 1.7786  0.9997  1.779 0.075226 .
relevel(trigen, ref = "nonbinary")male -1.1332   1.0222 -1.109 0.267607
freq_scaled                            -1.8204    0.5141 -3.541 0.000398 ***
sylb_scaled                            -0.7731    0.1383 -5.590 2.28e-08 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                     (Intr) rlvl(trgn,rf="nnbnry")f rlvl(trgn,rf="nnbnry")m
                         frq_sc
rlvl(trgn,rf="nnbnry")f -0.642
rlvl(trgn,rf="nnbnry")m -0.627 0.418
freq_scaled            0.158 -0.039                0.024
sylb_scaled            0.044 -0.038                0.015
    0.053
```

Same GLMER as above, but with default reference level for GENDER (female)

```
Generalized linear mixed model fit by maximum likelihood (Laplace
    Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: fricatedbi ~ trigen + freq_scaled + sylb_scaled + (1 | speaker/
    transcript) + (1 | text)
   Data: janvtvCscales2

     AIC      BIC   logLik deviance df.resid
  1228.7   1270.7   -606.3   1212.7     1403

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.0724 -0.3574 -0.1044 0.3203 6.5749

Random effects:
 Groups           Name       Variance Std.Dev.
 text             (Intercept) 5.289  2.300
 transcript:speaker (Intercept) 0.366 0.605
 speaker          (Intercept) 2.341  1.530
Number of obs: 1411, groups: text, 417; transcript:speaker, 74; speaker, 16

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.1306     0.7674   1.473 0.140670
trigenmale     -2.9117     1.0905  -2.670 0.007584 **
trigennonbinary -1.7784    0.9998  -1.779 0.075286 .
freq_scaled    -1.8204     0.5141  -3.541 0.000398 ***
sylb_scaled    -0.7731     0.1383  -5.590 2.27e-08 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) trgnml trgnnn frq_sc'
trigenmale -0.682
trignnnbnry -0.732 0.524
freq_scaled 0.090 0.058  0.039
sylb_scaled -0.010 0.049 0.038 0.053
```

Comparison of final model with a competing model with the gender term removed.

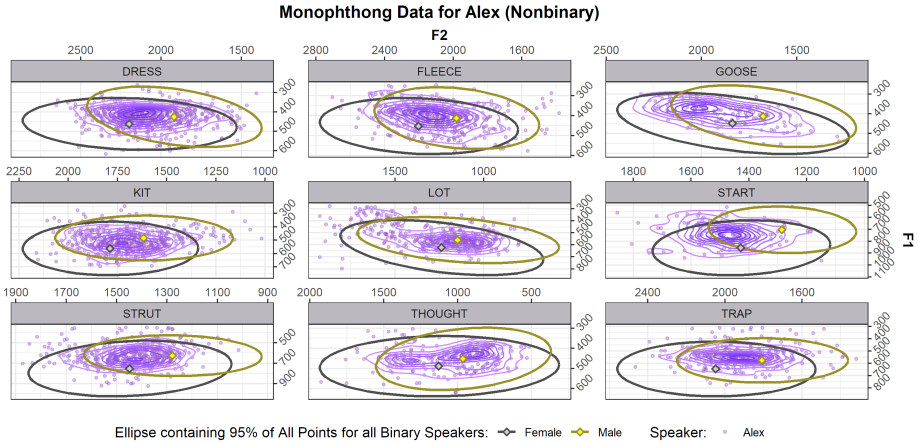
```
Data: relevelvtv
Models:
releveledModelnogen: fricatedbi ~ freq_scaled + sylb_scaled + (1 | speaker/
    transcript) +
releveledModelnogen:   (1 | text)
releveledModel: fricatedbi ~ relevel(trigen, ref = "nonbinary") +
    freq_scaled +
releveledModel:   sylb_scaled + (1 | speaker/transcript) + (1 | text)
                  Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
releveledModelnogen 6 1231.0 1262.5 -609.49 1219.0
releveledModel     8 1228.7 1270.7 -606.35 1212.7 6.2925    2    0.04301 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

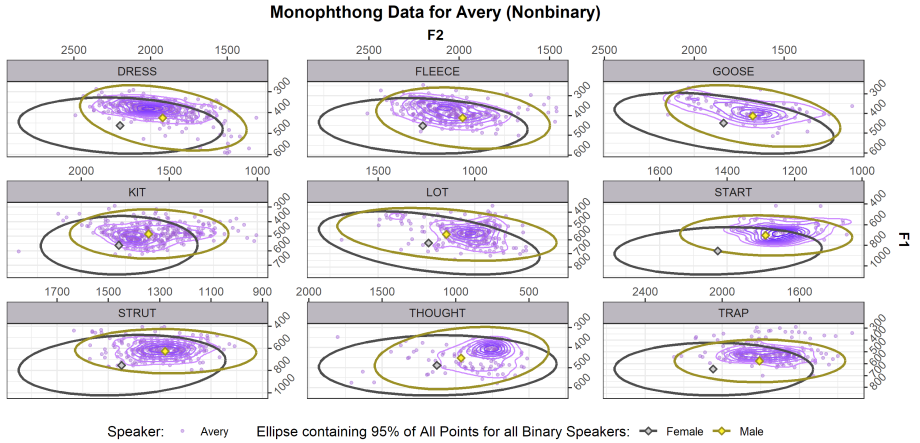
Appendix F

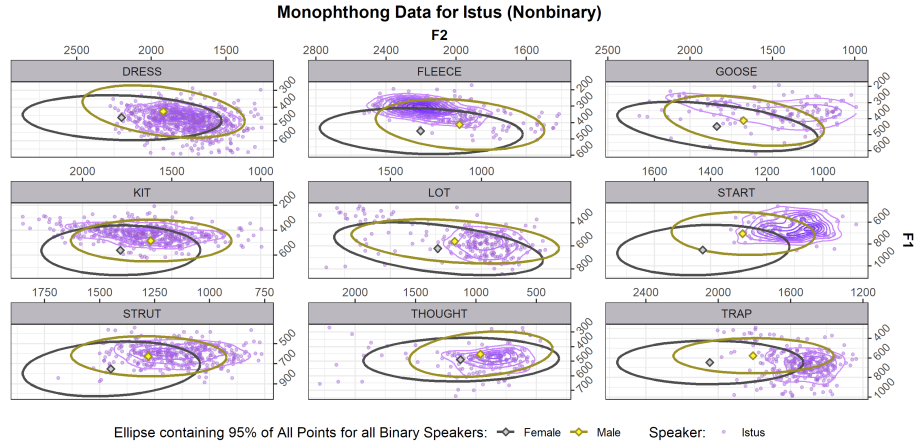
Full Monophthong Visualisations

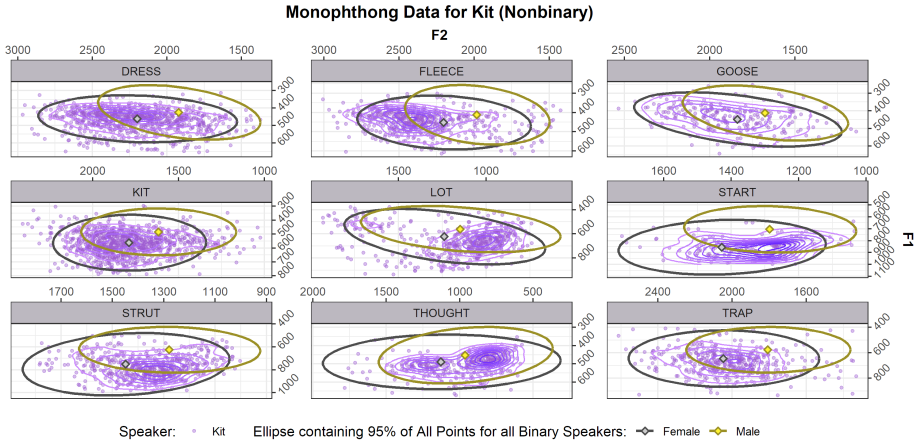
This Appendix contains visualisations of all monophthongs (excepting SCHWA), of all speakers (excepting Mello), with reference to the aggregated data of binary men and women. Points represent individual measurements per speaker, plotted alongside data ellipses representing 95% of all tokens for men and women. Contour lines are set to display 10 "bins" - approximately 10% of each speaker's tokens are within each contour. All monophthongs are shown, though only five show any statistically significant differences ($p < 0.05$) in F2 across gender (KIT, LOT, STRUT, TRAP, and THOUGHT, along F2 only).

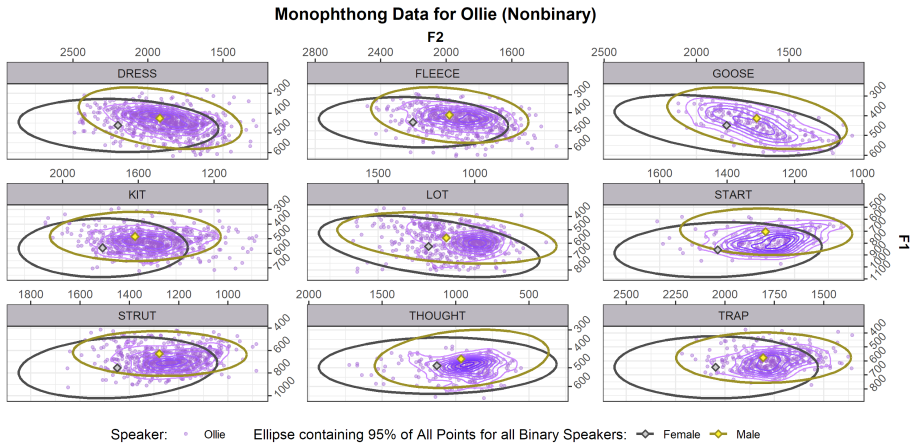
F.1 Nonbinary Speakers

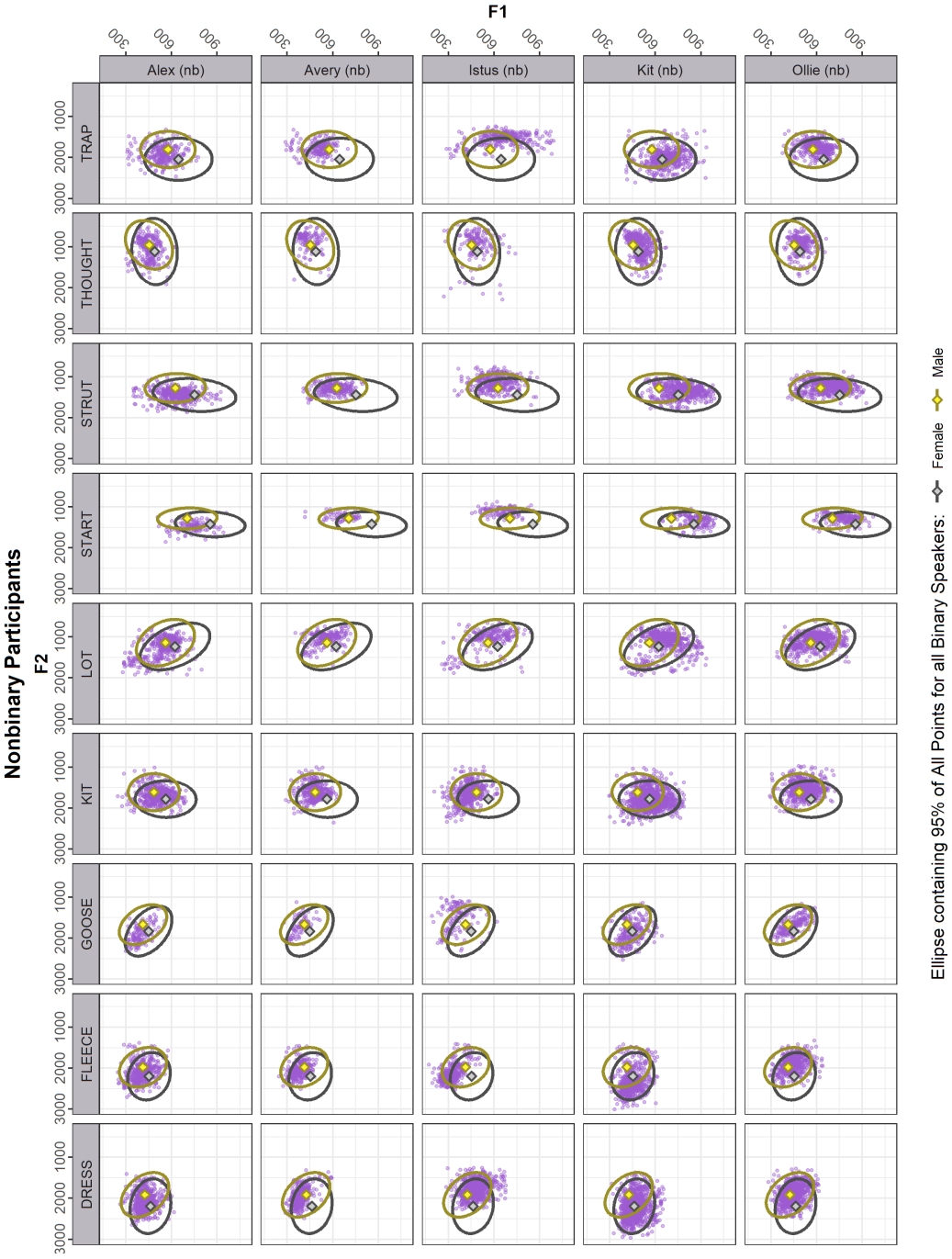




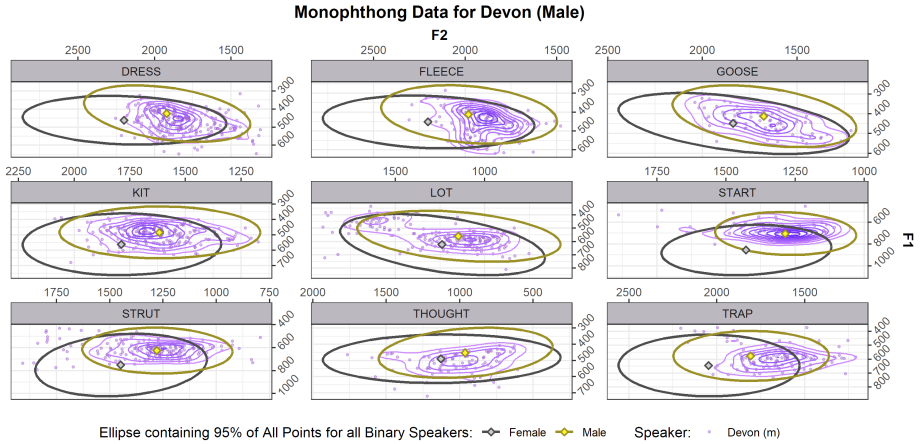

Monophthong Data for Istus (Nonbinary)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Ellipse containing 95% of All Points for all Binary Speakers: Female Male
Speaker: Istus

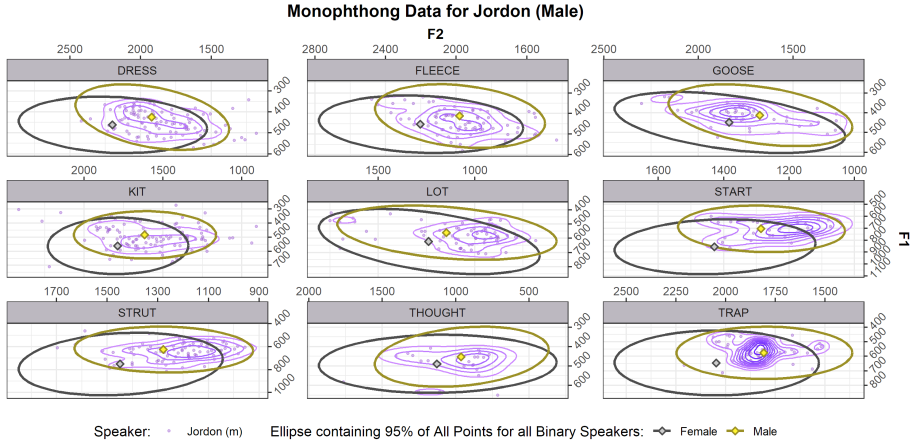

Monophthong Data for Kit (Nonbinary)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Kit
Ellipse containing 95% of All Points for all Binary Speakers: Female Male

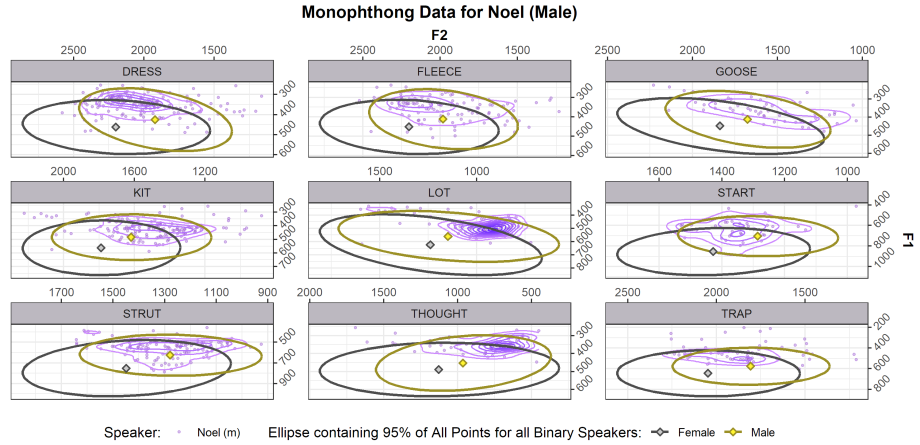

Monophthong Data for Ollie (Nonbinary)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Ollie
Ellipse containing 95% of All Points for all Binary Speakers: Female Male

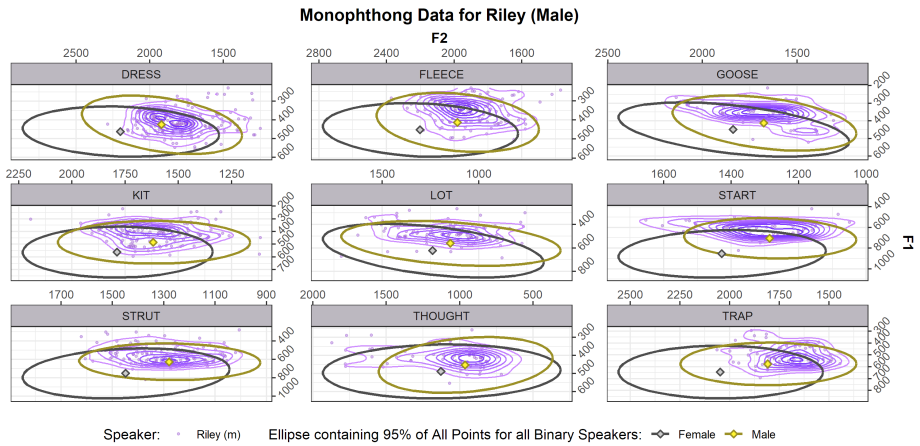

Nonbinary Participants
F1
F2
Alex (nb)
Avery (nb)
Istus (nb)
Kit (nb)
Ollie (nb)
TRAP
THOUGHT
STRUT
START
LOT
KIT
GOOSE
FLEECE
DRESS
300
600
900
1000
2000
3000
Ellipse containing 95% of All Points for all Binary Speakers:
Female
Male

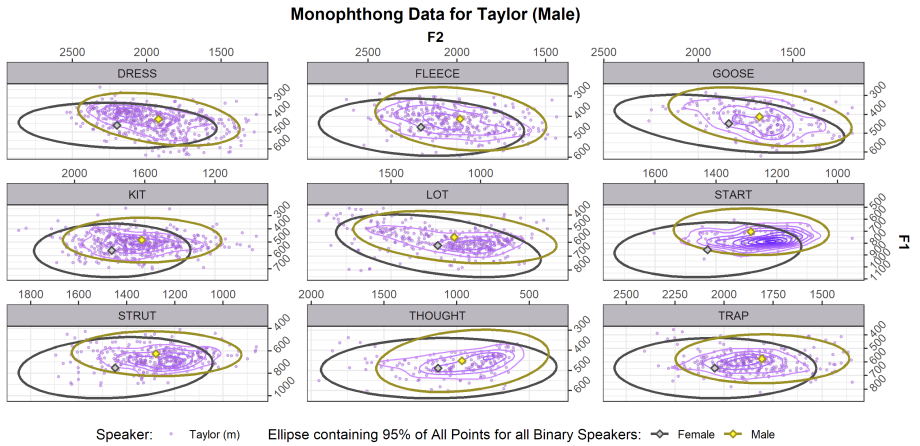
F.2 Binary Men

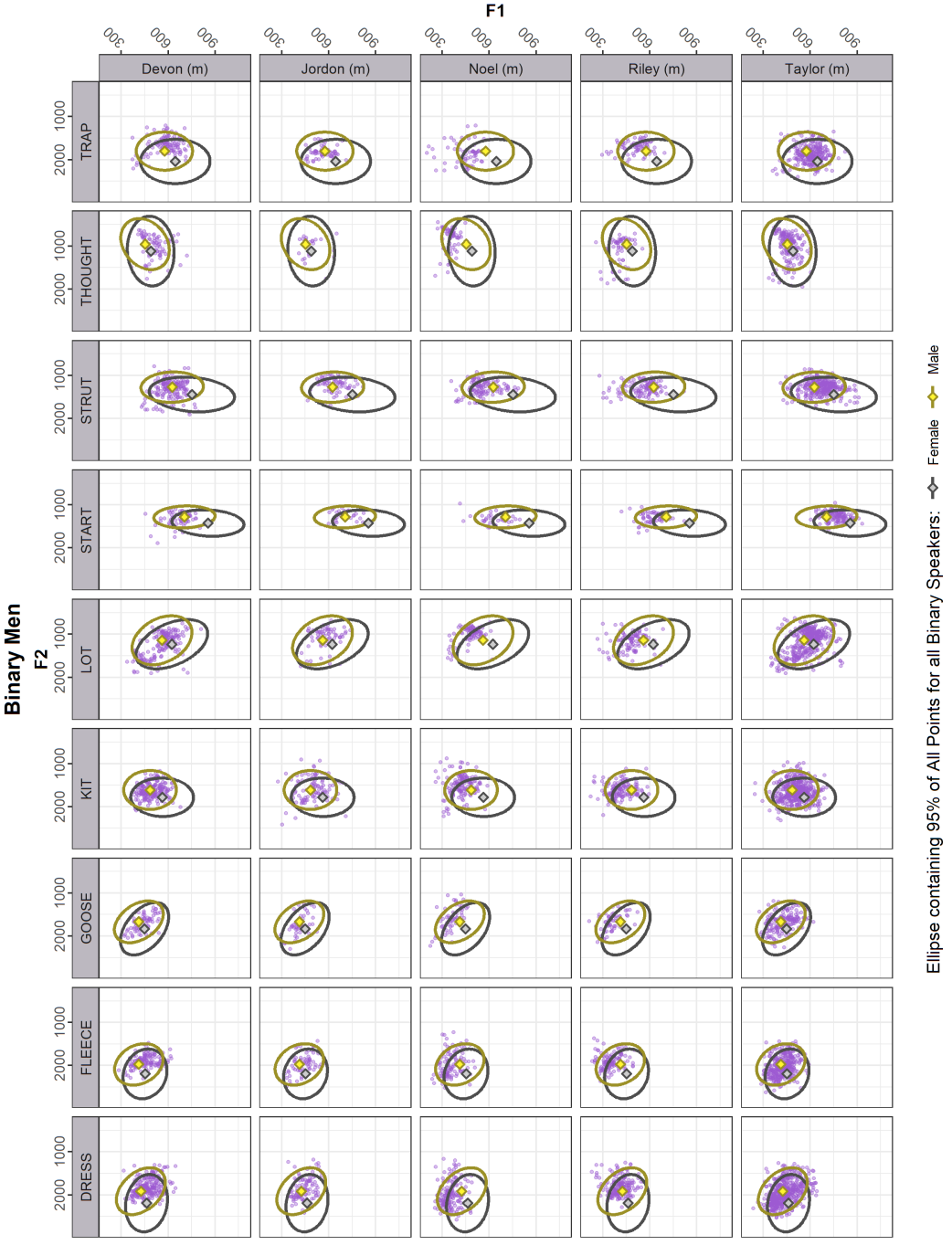




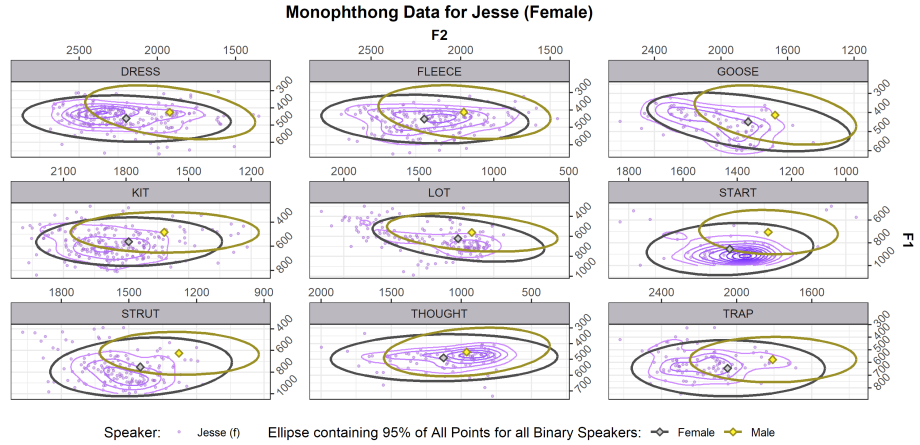

Monophthong Data for Noel (Male)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Noel (m)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male

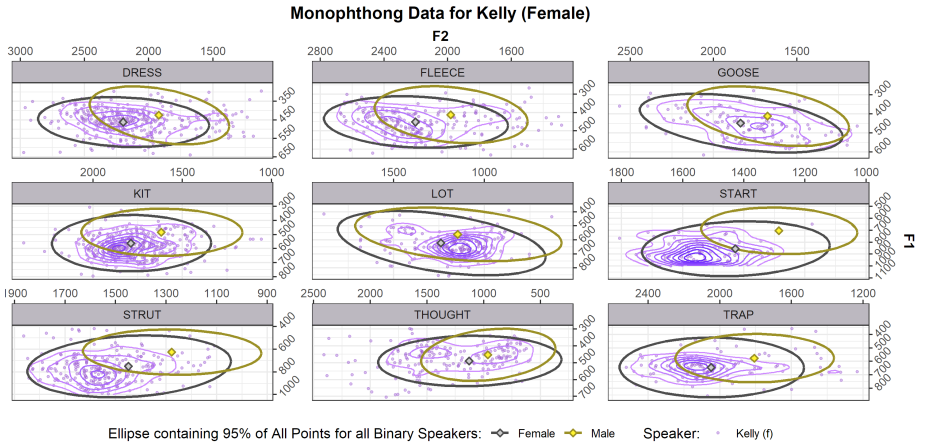

Monophthong Data for Riley (Male)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Riley (m)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male

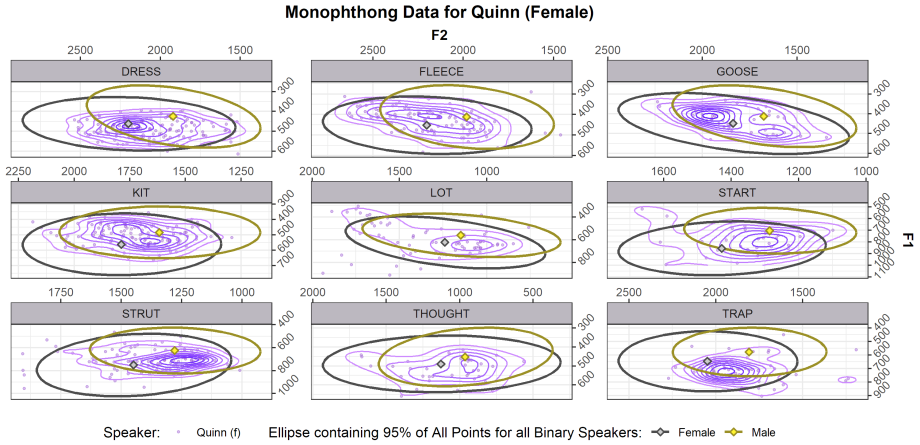

Monophthong Data for Taylor (Male)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Taylor (m)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male

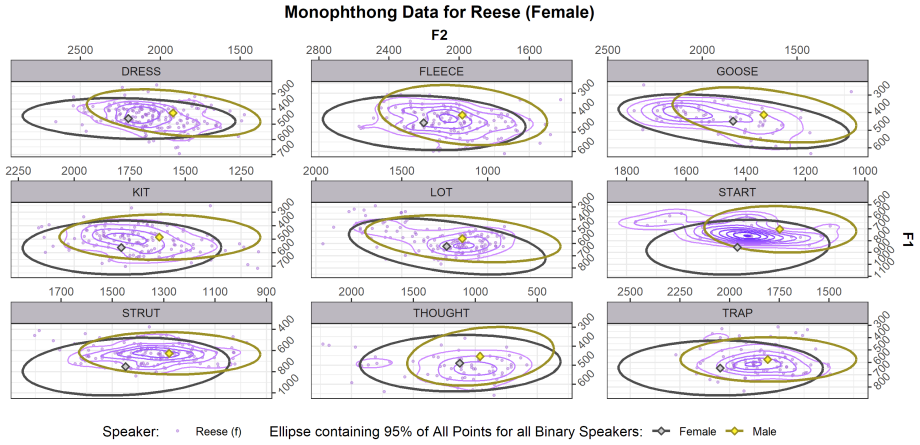

Binary Men
F1
F2
Devon (m)
Jordon (m)
Noel (m)
Riley (m)
Taylor (m)
TRAP
THOUGHT
STRUT
START
LOT
KIT
GOOSE
FLEECE
DRESS
300
600
900
1000
2000
Ellipse containing 95% of All Points for all Binary Speakers:
Female
Male

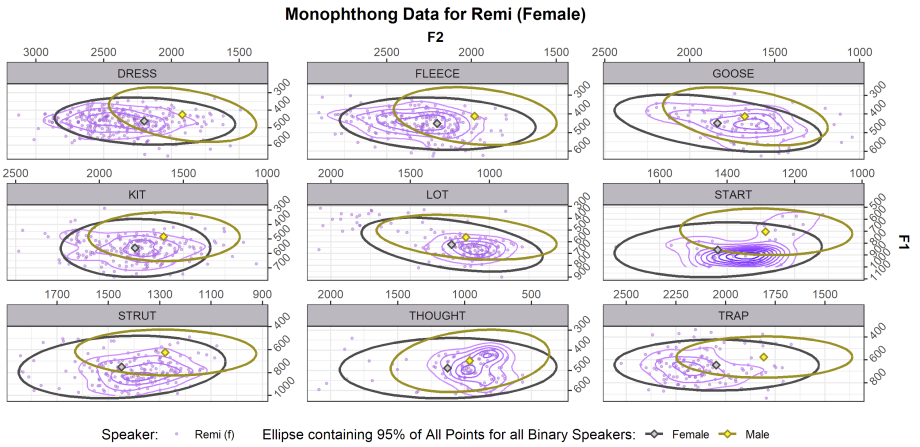
F.3 Binary Women

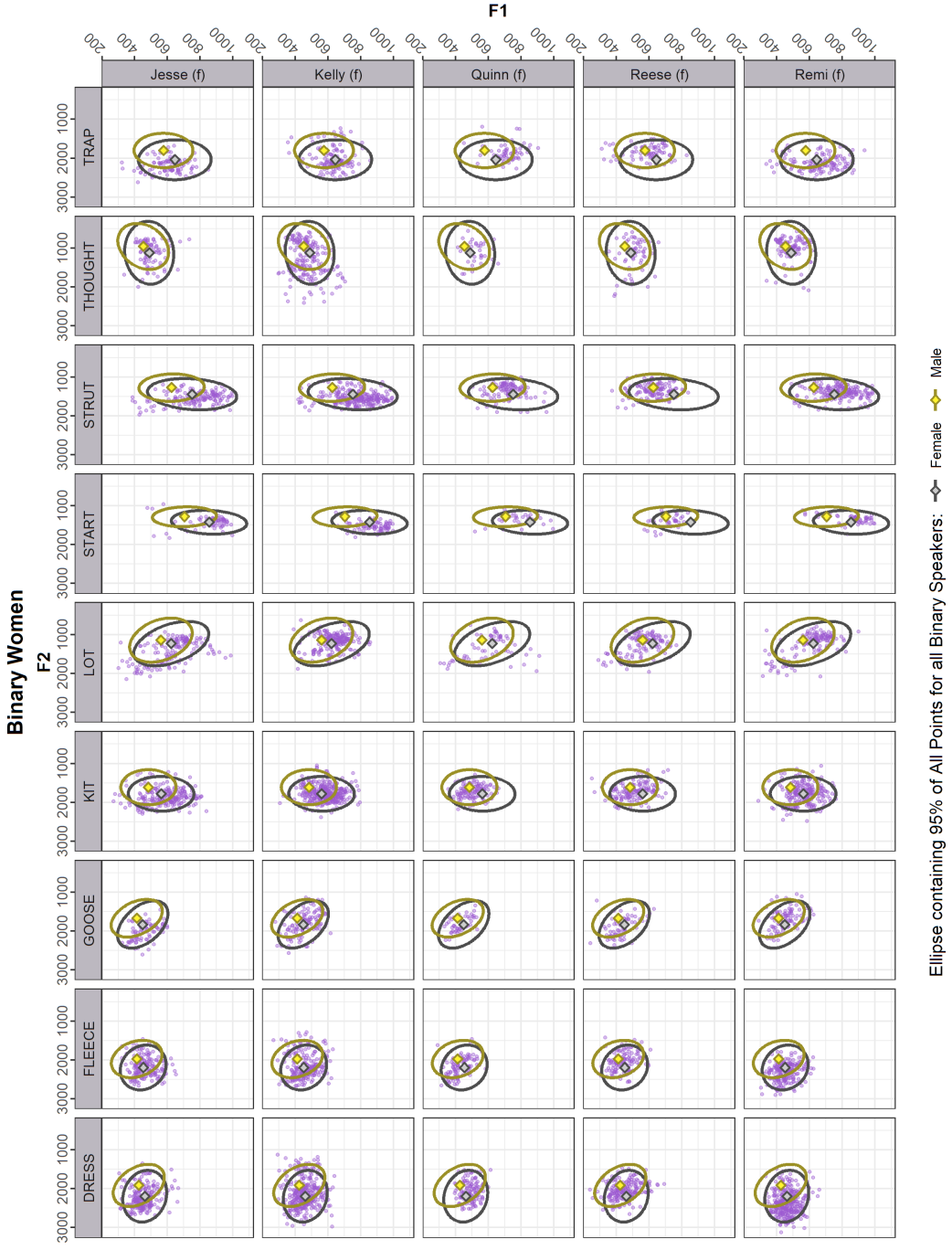





Monophthong Data for Quinn (Female)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Quinn (f)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male


Monophthong Data for Reese (Female)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Reese (f)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male


Monophthong Data for Remi (Female)
F2
F1
DRESS
FLEECE
GOOSE
KIT
LOT
START
STRUT
THOUGHT
TRAP
Speaker: Remi (f)
Ellipse containing 95% of All Points for all Binary Speakers: Female Male


Binary Women
F1
F2
Jesse (f)
Kelly (f)
Quinn (f)
Reese (f)
Remi (f)
TRAP
THOUGHT
STRUT
START
LOT
KIT
GOOSE
FLEECE
DRESS
Ellipse containing 95% of All Points for all Binary Speakers:
Female
Male

Appendix G

Intervocalic [t] Tagging Examples

This Appendix contains a selection of example spectrograms to illustrated the different categories of [t] productions. I have not included examples of productions coded as NA, such as ∅ and [ʔ]. I have also not provided an examples of [d], as they are very similar to the [ɾ] category, and were collapsed early in the tagging process.

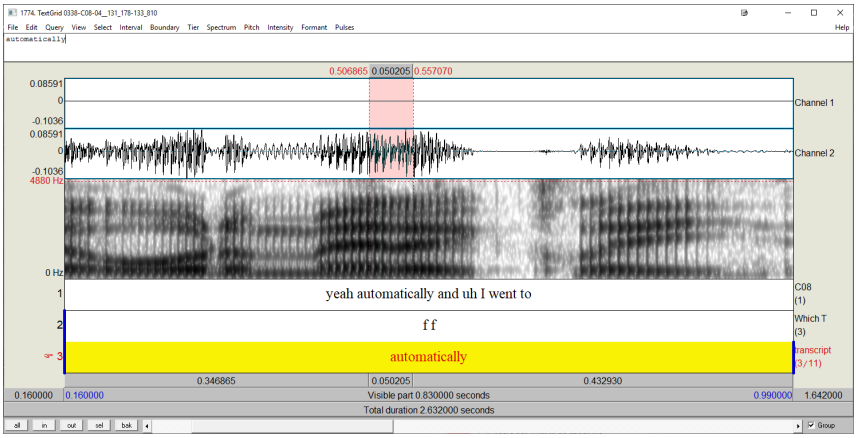


FIGURE G.1: TWO FLAPS
Two closures both coded as a flap (the second one is highlighted). Note the first shows a brief closure with a prototypical "pinch", while the second is barely visible in the spectrogram.
n = 642

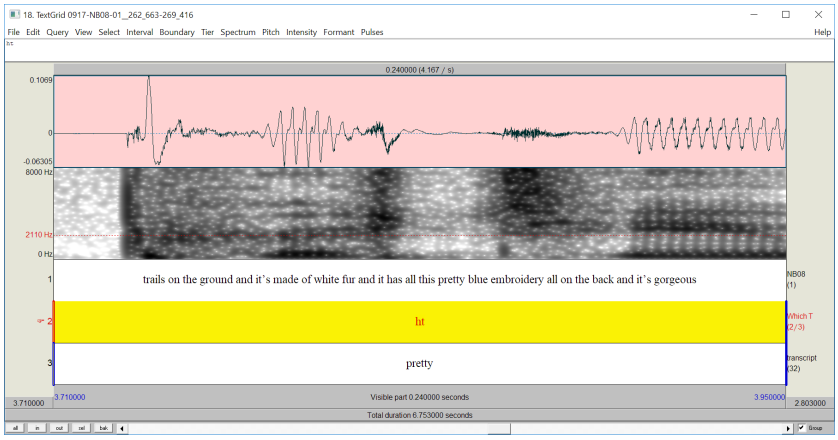


FIGURE G.2: PRE-ASPIRATION
Note the high-frequency noise on either side of the closure.
n = 41

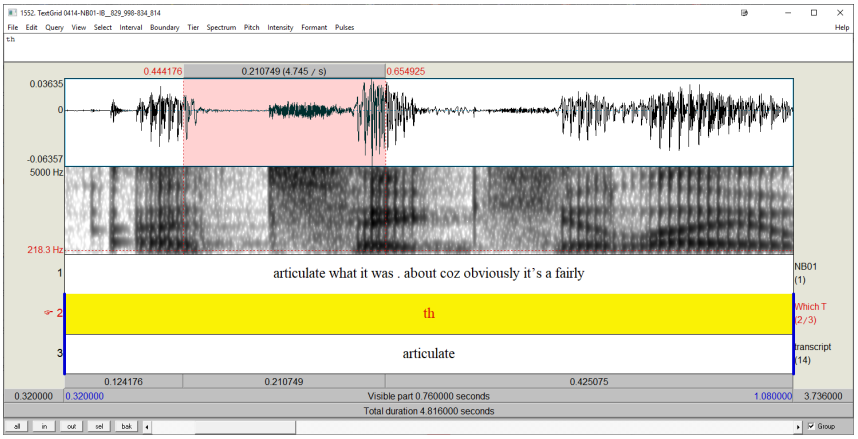


FIGURE G.3: CANONICAL [t^h]
N = 153

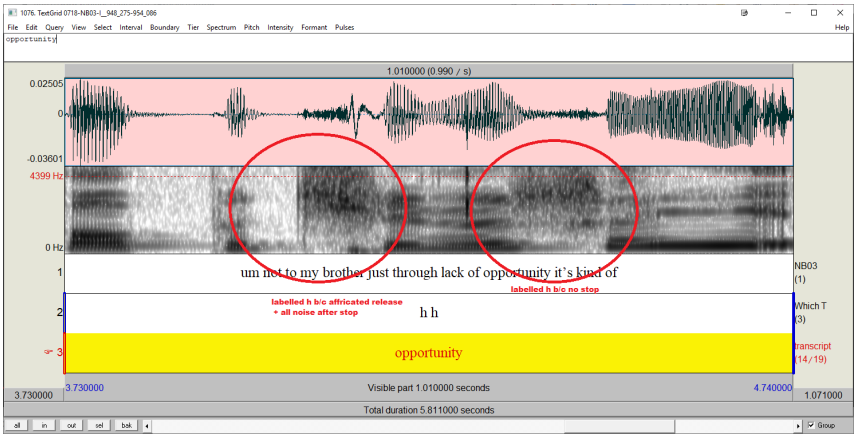


FIGURE G.4: AFFRICATES / FRICATIVES
n = 359

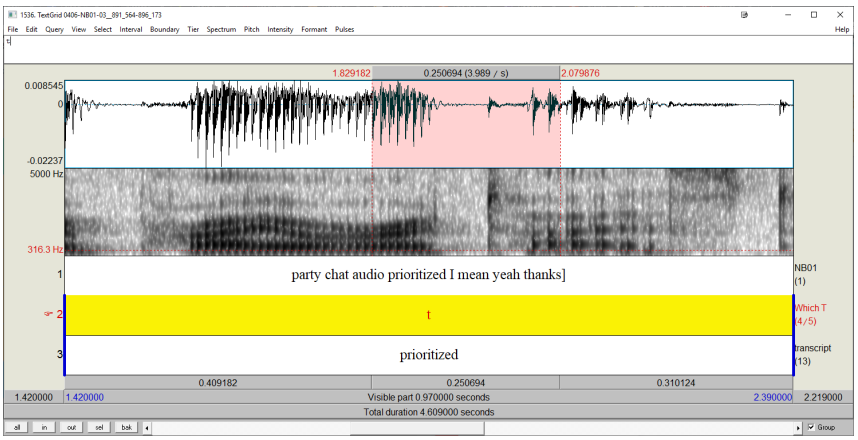
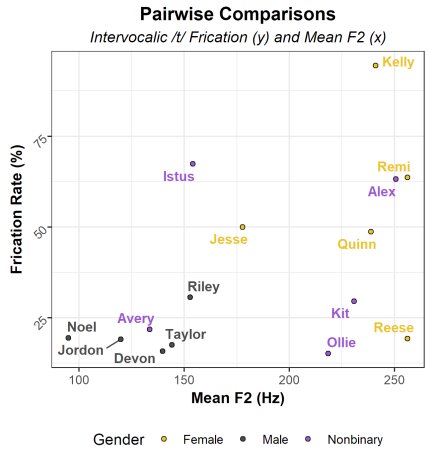


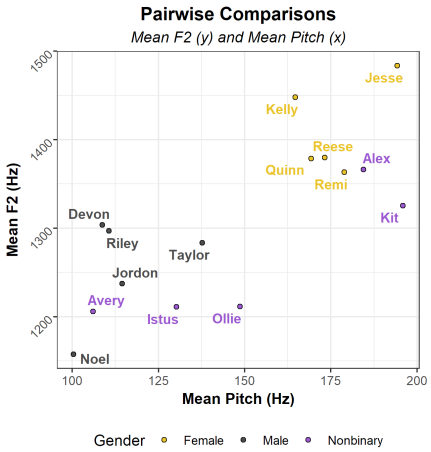
FIGURE G.5: CANONICAL [t]
No prevoicing, very little high-frequency noise and short duration compared to [t^h]
n = 17

Appendix H

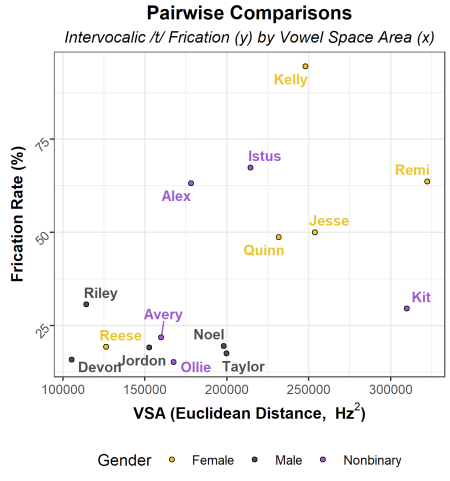
Complete Multiple Signal Visualisations



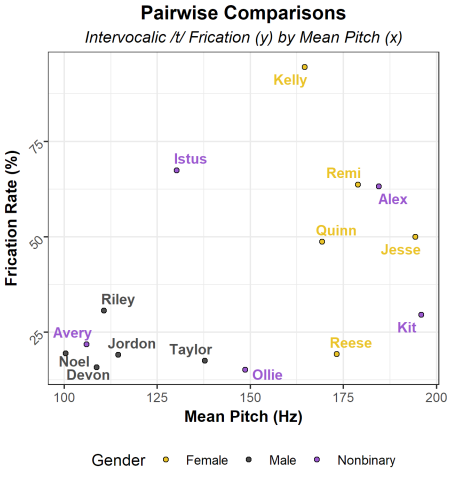
(i) Pairwise comparison of mean F2 of four monophthongs and intervocalic /t/ frication.



(ii) Pairwise comparison of mean pitch and mean F2 of four monophthongs.

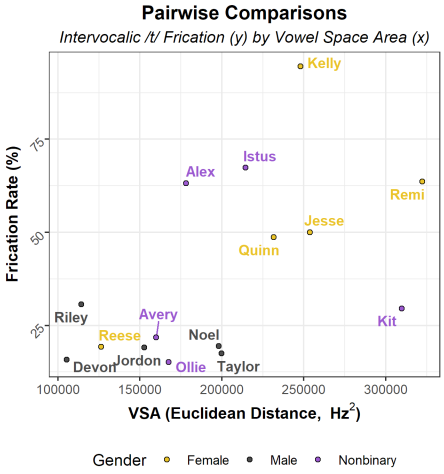


(iii) Pairwise comparison of mean pitch and VSA.

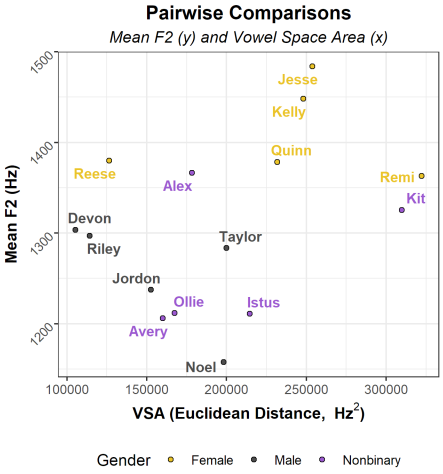


(iv) Pairwise comparison of mean pitch and intervocalic /t/ frication.

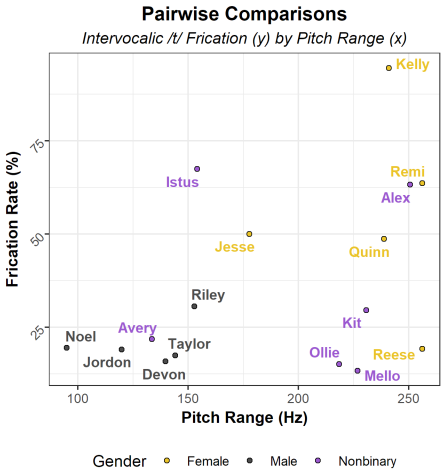
FIGURE H.1: Pairwise Comparisons: Multiple Signal Productions I



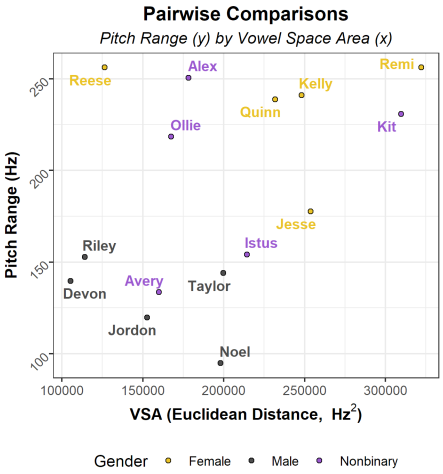
(i) Pairwise comparison of VSA and intervocalic /t/ frication.



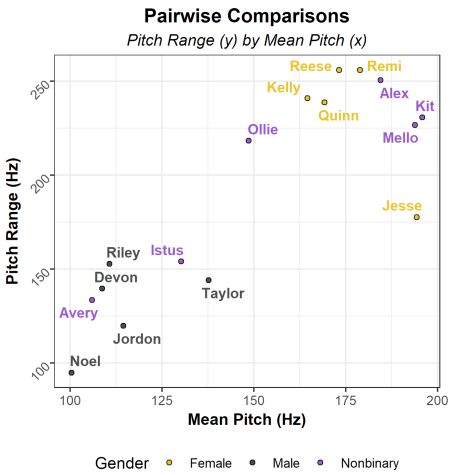
(ii) Pairwise comparison of VSA and mean F2 of four monophthongs.



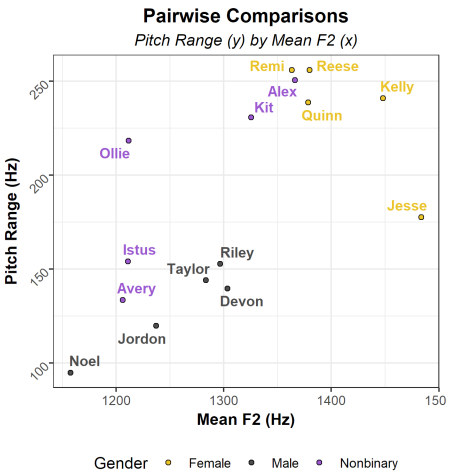
(iii) Pairwise comparison of pitch range and intervocalic /t/ frication.



(iv) Pairwise comparison of pitch range and VSA.

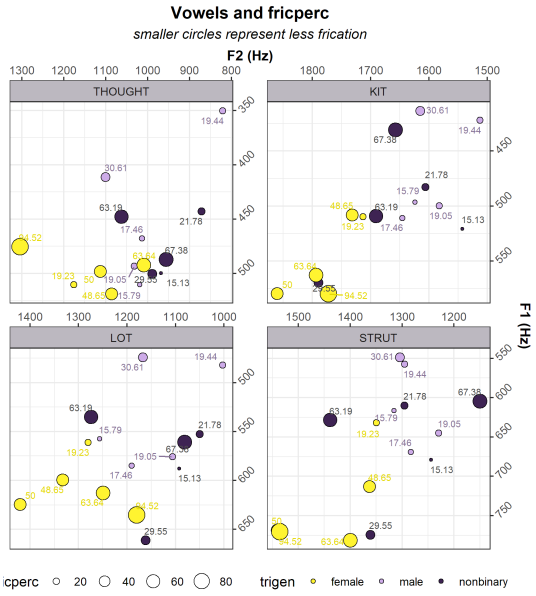


(v) Pairwise comparison of mean pitch and pitch range

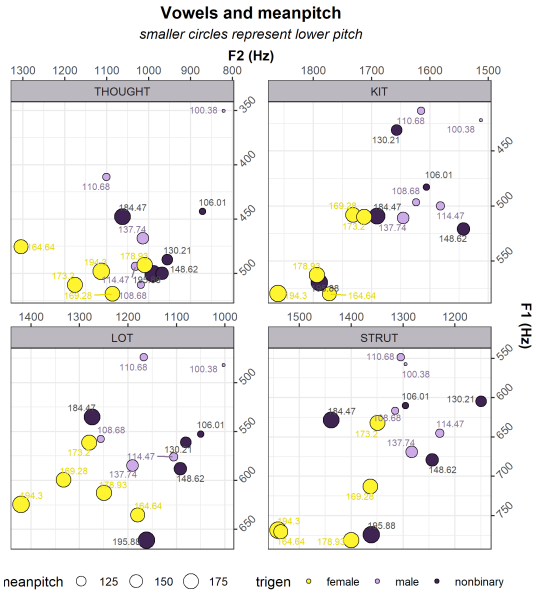


(vi) Pairwise comparison of pitch range and mean F2 of four monophthongs.

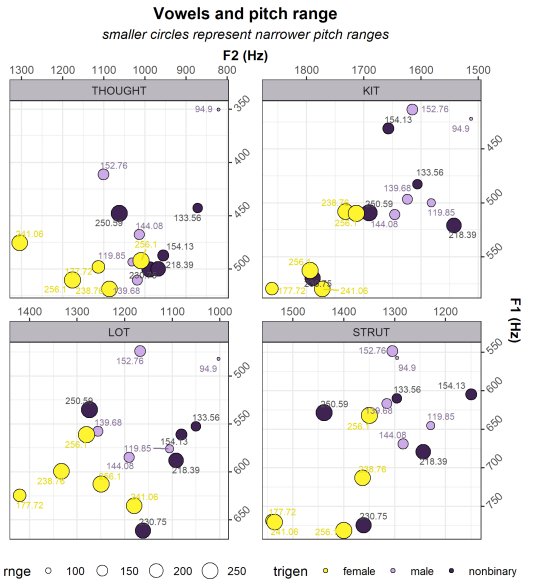
FIGURE H.2: Pairwise Comparisons: Multiple Signal Productions II



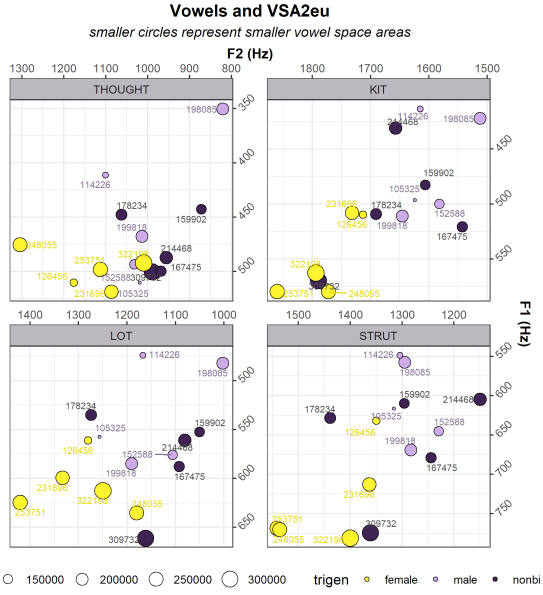
(i) vowels and frication



(ii) vowels and mean pitch



(iii) vowels and pitch range



(iv) vowels and vsa

FIGURE H.3: F1/F2 of four vowels and other variables

Bibliography

Abbou, Julie, Aron Arnold, Leann Brown, Maria Candea, James German, Tim Mahrt, and Oriana Reid-Collins (2018). "Gender and the voice: comparing metadiscourse in vlogs, interviews and scientific discourse". In: *Sociolinguistics Symposium 22*.

Adler, Richard (2007). "Peer reviewed article gender voice issues: voice and communication therapy for transsexual/transgender clients". In: *Voice and Speech Review* 5.1, pp. 293–299.

Agha, Asif (2003). "The social life of cultural value". In: *Language & communication* 23.3-4, pp. 231–273.

Ahl, Helene J (2002). "The making of the female entrepreneur: A discourse analysis of research texts on women's entrepreneurship". PhD thesis. Internationella Handelshögskolan.

Alarcón, José Antonio, Markus Bastir, and Antonio Rosas (2016). "Variation of mandibular sexual dimorphism across human facial patterns". In: *Homo* 67.3, pp. 188–202.

Ameijeiras-Alonso, Jose, Rosa M. Crujeiras, and Alberto Rodríguez-Casal (2021). "multimode: An R Package for Mode Assessment". In: *Journal of Statistical Software* 97.9, pp. 1–32. DOI: 10.18637/jss.v097.i09.

American Psychiatric Association (2013). "Gender Dysphoria". English. In: *Diagnostic and statistical manual of mental disorders : DSM-5*. 5th ed. American Psychiatric Association Arlington, VA. DOI: 10.1176/appi.books.9780890425596.dsm14.

Arai, Takayuki (2004). "History of chiba and kajiyama and their influence in modern speech science". In: *Proc. From Sound Sense* 50, pp. 115–120.

Armstrong, Marie M, Anthony J Lee, and David R Feinberg (2019). "A house of cards: bias in perception of body size mediates the relationship between voice pitch and perceptions of dominance". In: *Animal Behaviour* 147, pp. 43–51.

Arnold, Aron (2016). "Idéologies de genre et construction des savoirs en sciences phonétiques". In: *GLAD!. Revue sur le langage, le genre, les sexualités* 01.

Atkin, Albert (2022). "Peirce's Theory of Signs". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Fall 2022. Metaphysics Research Lab, Stanford University.

Bachorowski, Jo-Anne and Michael J Owren (1999). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech". In: *The Journal of the Acoustical Society of America* 106.2, pp. 1054–1063.

Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand (1991). "The apparent time construct". In: *Language variation and change* 3.3, pp. 241–264.

Baken, R. J. and Robert F. Orlikoff (2000). *Clinical measurement of speech and voice*. English. 2nd. San Diego: Singular Thomson Learning. ISBN: 1565938690;9781565938694;

Barbee, Harry and Douglas Schrock (2019). "Un/Gendering social selves: How nonbinary people navigate and experience a binarily gendered world". In: *Sociological Forum*. Vol. 34. 3. Wiley Online Library, pp. 572–593.

Barreda, Santiago (2020). "Vowel normalization as perceptual constancy". In: *Language* 96.2, pp. 224–254.

Barreda, Santiago (2021). "Perceptual validation of vowel normalization methods for variationist research". In: *Language Variation and Change*, pp. 1–27.

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck, Peter Green, and Maintainer Ben Bolker (2015). "Package 'lme4'". In: *Convergence* 12.1, p. 2.

Bauer, Laurie, Paul Warren, Dianne Bardsley, Marianna Kennedy, and George Major (2007). "New zealand english". In: *Journal of the International Phonetic Association* 37.1, pp. 97–102.

Bayard, Donn (1987). "Class and change in New Zealand English: a summary report". In: *Te Reo* 30, pp. 3–36.

Becker, Kara, S Khan, and Lal Zimman (2015). "Creaky voice in a diverse gender sample: Challenging ideologies about sex, gender and creak in American English". In: *NWAV44 Toronto*.

Bell, Allan (1977). "The language of radio news in Auckland: A sociolinguistic study of style, audience & subediting variation". PhD thesis. University of Auckland.

— (1984). "Language style as audience design". In: *Language in society* 13.2, pp. 145–204.

— (2016). "18 Succeeding waves: Seeking sociolinguistic theory for the twenty-first century". In: *Sociolinguistics: Theoretical Debates*, p. 391.

Berberick, Stephanie N et al. (2018). "The Paradox of Trans Visibility: Interrogating the "Year of Trans Visibility"". In: *Journal of Media Critiques* 4.13, pp. 123–144.

Berger, Jonah and Morgan Ward (2010). "Subtle signals of inconspicuous consumption". In: *Journal of Consumer Research* 37.4, pp. 555–569.

Bishop, Jason and Patricia Keating (2012). "Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex". In: *The Journal of the Acoustical Society of America* 132.2, pp. 1100–1112.

Boë, Louis-Jean, Jean-Louis Heim, Kiyoshi Honda, Shinji Maeda, Pierre Badin, and Christian Abry (2007). "The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a)". In: *Journal of Phonetics* 35.4, pp. 564–581.

Boersma, Paul and David Weenink (Jan. 2016). "PRAAT: Doing phonetics by computer (Version 6.0.22)". In: *https://www.praat.org/*.

Boersma, Paul et al. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". In: *Proceedings of the institute of phonetic sciences*. Vol. 17. 1193. Citeseer, pp. 97–110.

Bowling, Daniel L, Maxime Garcia, Jacob C Dunn, R Ruprecht, Asha Stewart, K-H Frommolt, and WT Fitch (2017). "Body size and vocalization in primates and carnivores". In: *Scientific reports* 7.1, pp. 1–11.

Bradlow, Ann R, Gina M Torretta, and David B Pisoni (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics". In: *Speech communication* 20.3, p. 255.

Brand, James, Jennifer Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy (2019). "Systematic covariation of monophthongs across speakers of New Zealand English". In: *Proceedings of the 19th International Congress of Phonetic Sciences*. Ed. by S. Calhoun, P. Escudero, M. Tabain, and P. Warren. Vol. 10, pp. 1878–1881.

Brand, James, Jennifer Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy (forthcoming). "At the margins of vowel-space change: Leaders and laggers in the evolution of vocalic subsystems". In: *Journal of Phonetics*.

Britto, Anthea I and Philip C Doyle (1990). "A comparison of habitual and derived optimal voice fundamental frequency values in normal young adult speakers". In: *Journal of Speech and Hearing Disorders* 55.3, pp. 476–484.

Brumm, Henrik and Sue Anne Zollinger (2011). "The evolution of the Lombard effect: 100 years of psychoacoustic research". In: *Behaviour* 148.11-13, pp. 1173–1198.

Bucholtz, Mary and Kira Hall (2004). "Language and identity". In: *A companion to linguistic anthropology* 1, pp. 369–394.

— (2005). "Identity and interaction: A sociocultural linguistic approach". In: *Discourse studies* 7.4-5, pp. 585–614.

Burman, Prabir and Wolfgang Polonik (2009). "Multivariate mode hunting: Data analytic tools with measures of significance". In: *Journal of Multivariate Analysis* 100.6, pp. 1198–1218.

Burns, Ericka (2021). "Now you're speaking my language: Linguistic trends of Black queer youth". In: *Lavender Languages and Linguistics 27*.

Butler, Judith (1990). "Gender trouble: Feminism and the subversion of identity". In.

— (2011). *Bodies that matter: On the discursive limits of sex*. Routledge London.

Butler, Judith et al. (2004). *Undoing gender*. Routledge New York.

Calás, Marta B and Linda Smircich (1996). "From the 'woman's point of view': Feminist approaches to organization studies". In: *Handbook of organization studies*. Ed. by Stewart R Clegg, Cynthia Hardy, and Walter R Nord. London: SAGE publications Ltd, pp. 218–57.

Calder, Jeremy (2020). "Language and Sexuality: Language and LGBTQ+ Communities". In: *The International Encyclopedia of Linguistic Anthropology*, pp. 1–7.

Cameron, Deborah and Don Kulick (2003). *Language and sexuality*. Cambridge University Press.

Carroll, Rachel (2020). "Making whiteness visible: transgender, race and the paradoxes of in/visibility in Orlando (1928), The Passion of New Eve (1977) and Sacred Country (1992)". In: *European Journal of English Studies* 24.1, pp. 13–24.

Cavanagh, Sheila L (2010). *Queering bathrooms: Gender, sexuality, and the hygienic imagination*. University of Toronto Press.

Chandler, Daniel and Rod Munday (2011). *biological essentialism*. DOI: 10.1093/acref/9780199568758.013.0218. URL: https://www.oxfordreference.com/view/10.1093/acref/9780199568758.001.0001/acref-9780199568758-e-218.

Charlton, Benjamin D, William AH Ellis, Allan J McKinnon, Gary J Cowin, Jacqui Brumm, Karen Nilsson, and W Tecumseh Fitch (2011). "Cues to body size in the formant spacing of male koala (Phascolarctos cinereus) bellows: honesty in an exaggerated trait". In: *Journal of Experimental Biology* 214.20, pp. 3414–3422.

Chela-Flores, Godsuno (2006). "A Look at a Forgotten Dimension in Dialectal Description: Articulatory Settings or Corpo-Vocal Memory?" In: *Dialectologia et Geolinguistica: Journal of the International Society for Dialectology and Geolinguistics* 2006.14, pp. 3–11. DOI: doi:10.1515/dig.2006.2006.14.3. URL: https://doi.org/10.1515/dig.2006.2006.14.3.

Chiba, Tsutomu and Masato Kajiyama (1958). *The vowel: Its nature and structure*. Vol. 652. Phonetic society of Japan Tokyo.

Childers, Donald G and Ke Wu (1991). "Gender recognition from speech. Part II: Fine analysis". In: *The Journal of the Acoustical society of America* 90.4, pp. 1841–1856.

Clark, Beth A, Jaimie F Veale, Marria Townsend, Hélène Frohard-Dourlent, and Elizabeth Saewyc (2018). "Non-binary youth: Access to gender-affirming primary health care". In: *International Journal of Transgenderism* 19.2, pp. 158–169.

Clark, Lynn, Helen MacGougan, Jennifer Hay, and Liam Walsh (2016). ""Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus". In: *Ampersand* 3, pp. 13–20.

Cler, Gabriel J, Victoria S McKenna, Kimberly L Dahl, and Cara E Stepp (2020). "Longitudinal case study of transgender voice changes under testosterone hormone therapy". In: *Journal of Voice* 34.5, pp. 748–762.

Coats, Susan and Eliot R Smith (1999). "Perceptions of gender subtypes: Sensitivity to recent exemplar activation and in-group/out-group differences". In: *Personality and Social Psychology Bulletin* 25.4, pp. 516–526.

Coleman, Ralph O (1976). "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice". In: *Journal of speech and hearing research* 19.1, pp. 168–180.

Cordoba, Sebastian (2020). "Exploring non-binary genders: Language and identity". PhD thesis. De Montfort University.

Corwin, Anna I (2009). "Language and gender variance: Constructing gender beyond the male/female binary". In: *Electronic Journal of Human Sexuality* 12.4.

— (2017). "Emerging genders: semiotic agency and the performance of gender among genderqueer individuals." In: *Gender & Language* 11.2.

Costello, Cary Gabriel (2016). "Intersex and trans* communities: Commonalities and tensions". In: *Transgender and Intersex: theoretical, practical, and artistic perspectives*. Springer, pp. 83–113.

Coupland, Nikolas (1980). "Style-shifting in a Cardiff work-setting". In: *Language in society* 9.1, pp. 1–12.

Coupland, Nikolas and Hywel Bishop (2007). "Ideologised values for British accents 1". In: *Journal of sociolinguistics* 11.1, pp. 74–93.

Crenshaw, Kimberlé (1989). "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics". In: *u. Chi. Legal f.*, p. 139.

— (2017). *On intersectionality: Essential writings*. The New Press.

Cupples, Julie, Victoria Guyatt, and Jamie Pearce (2007). ""Put on a Jacket, You Wuss": Cultural Identities, Home Heating, and Air Pollution in Christchurch, New Zealand". In: *Environment and Planning A: Economy and Space* 39.12, pp. 2883–2898. DOI: 10.1068/a3932. eprint: https://doi.org/10.1068/a3932. URL: https://doi.org/10.1068/a3932.

Dahiru, Tukur (2008). "P-value, a true test of statistical significance? A cautionary note". In: *Annals of Ibadan postgraduate medicine* 6.1, pp. 21–26.

Dallaston, Katherine and Gerard Docherty (2019). "Estimating the prevalence of creaky voice: A fundamental frequency-based approach". In: *Proceedings of the 19th International Congress of Phonetic Sciences. Australasian Speech Science and Technology Association Inc*, pp. 532–536.

Davidson, Lisa (2019). "The effects of pitch, gender, and prosodic context on the identification of creaky voice". In: *Phonetica* 76.4, pp. 235–262.

— (2021). "The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 12.3, e1547.

Davies, Shelagh and Joshua M Goldberg (2006). "Clinical aspects of transgender speech feminization and masculinization". In: *International Journal of Transgenderism* 9.3-4, pp. 167–196.

Davis, Julie Hirschfeld and Matt Apuzzo (May 12, 2016). "U.S. Directs Public Schools to Allow Transgender Access to Restrooms". In: *The New York Times*. URL: https://www.nytimes.com/2016/05/13/us/politics/obama-administration-to-issue-decree-on-transgender-access-to-school-restrooms.html (visited on 09/16/2019).

Davis, Pamela J, Shi Ping Zhang, Alison Winkworth, and Richard Bandler (1996). "Neural control of vocalization: Respiratory and emotional influences". In: *Journal of Voice* 10.1, pp. 23–38.

De Beauvoir, Simone (1956). *The second sex*. Jonathan Cape.

De Boer, Bart (2010). "Modelling vocal anatomy's significant effect on speech". In: *Journal of Evolutionary Psychology* 8.4, pp. 351–366.

Dea, Shannon (2016). *Beyond the binary: Thinking about sex and gender*. Broadview Press.

Delahunt, John W, Hayley J Denison, Dalice A Sim, Jemima J Bullock, and Jeremy D Krebs (2018). "Increasing rates of people identifying as transgender presenting to Endocrine Services in the Wellington region". In: *New Zealand Medical Journal* 131.1468, pp. 33–42.

Derrick, Donald and Bryan Gick (2021). "Gait change in tongue movement". In: *Scientific Reports* 11.1, pp. 1–14.

Diehl, Randy L, Björn Lindblom, Kathryn A Hoemeke, and Richard P Fahey (1996). "On explaining certain male-female differences in the phonetic realization of vowel categories". In: *Journal of phonetics* 24.2, pp. 187–208.

Dipzinski, Mitchel and Natasha Abner (2021). "A Spoken-Word Poet's Gestural Guide to Creating a Queer World". In: *Lavender Languages and Linguistics 27*.

Docherty, Gerry, Jennifer Hay, and Abby Walker (2006). "Sociophonetic patterning of phrase-final /t/ in New Zealand English". In: *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, pp. 378–383.

D'Onofrio, Annette, Teresa Pratt, and Janneke Van Hofwegen (2019). "Compression in the California Vowel Shift: Tracking generational sound change in California's Central Valley". In: *Language Variation and Change* 31.2, pp. 193–217.

Dorreen, Kieran (2017). "Fundamental frequency distributions of bilingual speakers in forensic speaker comparison". MA thesis. University of Canterbury.

Drager, Katie (2011). "Speaker age and vowel perception". In: *Language and Speech* 54.1, pp. 99–121.

Drager, Katie K (2015). *Linguistic variation, identity construction and cognition*. Language Science Press.

Eckert, Penelope (1989). "The whole woman: Sex and gender differences in variation". In: *Language variation and change* 1.3, pp. 245–267.

— (2003). "The meaning of style". In: *Proceedings of the Eleventh Annual Symposium about Language and Society, Austin*.

— (2008). "Variation and the indexical field". In: *Journal of sociolinguistics* 12.4, pp. 453–476.

— (2012). "Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation". In: *Annual review of Anthropology* 41, pp. 87–100.

— (2019). "The limits of meaning: Social indexicality, variation, and the cline of interiority". In: *Language* 95.4, pp. 751–776.

— (2021). "Gender and the Third Wave of variation study." In: *Gender & Language* 15.2.

Eckert, Penelope and Sally McConnell-Ginet (1992). "Communities of practice: Where language, gender and power all live". In: *Locating power: Proceedings of the second Berkeley women and language conference*. Vol. 1. Berkeley, CA: Berkeley University, pp. 89–99.

Eckert, Penelope and Robert J Podesva (2021). "Non-binary approaches to gender and sexuality". In: *The Routledge Handbook of Language, Gender, and Sexuality*, p. 2.

Emma, Caitlin (May 12, 2016). "Obama administration releases directive on transgender rights to school bathrooms". In: *Politico*. URL: https://www.politico.

com/story/2016/05/obama-administration-title-ix-transgender-student-rights-223149 (visited on 09/16/2019).

Engels, Friedrich (2021). *The origin of the family, private property and the state*. Verso Books.

Evans, Betsy E (2013). ""Everybody sounds the same": Otherwise overlooked ideology in perceptual dialectology". In: *American speech* 88.1, pp. 63–80.

Evans, Sarah, Nick Neave, and Delia Wakelin (2006). "Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice". In: *Biological psychology* 72.2, pp. 160–163.

Fant, Gunnar (1950a). *Transmission Properties of the Vocal Tract, Quarterly Progress Report July-September*. Tech. rep. M.I.T. Acoustics Laboratory.

— (1950b). *Transmission Properties of the Vocal Tract, Quarterly Progress Report October-December*. Tech. rep. M.I.T. Acoustics Laboratory.

— (1951). "Transmission Properties of the Vocal Tract—A Study Based on Analysis of Swedish Vowels and Consonants". In: *The Journal of the Acoustical Society of America* 23.1, pp. 148–148.

— (1952). *Transmission Properties of the Vocal Tract, Technical Report*. Tech. rep. 12. M.I.T. Acoustics Laboratory.

— (1966). "A note on vocal tract size factors and non-uniform F-pattern scalings". In: *Speech Transmission Laboratory Quarterly Progress and Status Report* 1, pp. 22–30.

— (1975). "Non-uniform vowel normalization". In: *STL-QPSR* 16.2-3, pp. 1–19.

Fiasson, Romain et al. (2016). "Frication, pre-aspiration and tapping of medial/t/in New Zealand English". In: *Te Reo* 59, pp. 47–69.

Fitch, W Tecumseh (1997). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques". In: *The Journal of the Acoustical Society of America* 102.2, pp. 1213–1222.

Fitch, W Tecumseh and Jay Giedd (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging". In: *The Journal of the Acoustical Society of America* 106.3, pp. 1511–1522.

Fletcher, Annalise R, Megan J McAuliffe, Kaitlin L Lansford, and Julie M Liss (2015). "The relationship between speech segment duration and vowel centralization in a group of older speakers". In: *The Journal of the Acoustical Society of America* 138.4, pp. 2132–2139.

Flynn, Nicholas (2011). "Comparing vowel formant normalisation procedures". In: *York papers in linguistics* 2.11, pp. 1–28.

Foucault, Michel (2008). *The History of Sexuality, Volume I*.

Fox, Robert A and Ewa Jacewicz (2008). "Analysis of total vowel space areas in three regional dialects of American English". In: *Journal of the Acoustical Society of America* 123.5, p. 3068.

Fox, Robert Allen and Ewa Jacewicz (2017). "Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English". In: *The Journal of the Acoustical Society of America* 142.1, pp. 444–459.

Freeman, Alan (May 18, 2016). "A new transgender rights bill may take Canada in a different direction". In: *The Washington Post*. URL: https://www.washingtonpost.com/news/worldviews/wp/2016/05/17/keeping-campaign-promise-justin-trudeau-moves-to-protect-transgender-canadians/ (visited on 10/15/2019).

Fromont, Robert and Jennifer Hay (2012). "LaBB-CAT: An annotation store". In: *Proceedings of the Australasian Language Technology Association Workshop 2012*, pp. 113–117.

Gamba, Marco, Olivier Friard, and Cristina Giacoma (2012). "Vocal tract morphology determines species-specific features in vocal signals of lemurs (Eulemur)". In: *International Journal of Primatology* 33.6, pp. 1453–1466.

Garmpi, Angel (2021). "Construction of Non-Binary Identities in Narrative Discourse". In: *Lavender Languages and Linguistics 27*.

Garnier, Maëva and Nathalie Henrich (2014). "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?" In: *Computer Speech & Language* 28.2, pp. 580–597.

Gelfer, Marylou Pausewang and Victoria A Mikos (2005). "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels". In: *Journal of Voice* 19.4, pp. 544–554.

Gelfer, Marylou Pausewang and Kevin J Schofield (2000). "Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male". In: *Journal of Voice* 14.1, pp. 22–33.

Gherardi, Silvia (1995). *Gender, symbolism and organizational cultures*. Sage.

Gick, Bryan, Ian Wilson, Karsten Koch, and Clare Cook (2004). "Language-specific articulatory settings: Evidence from inter-utterance rest position". In: *Phonetica* 61.4, pp. 220–233.

Giroux, Henry A and David E Purpel (1983). *The hidden curriculum and moral education: Deception or discovery?* McCutchan Publishing Corporation.

Gnevsheva, Ksenia (2015). "Acoustic analysis in the Accents of Non-Native English (ANNE) corpus". In: *International Journal of Learner Corpus Research* 1.2, pp. 256–267.

Goldstein, Ursula Gisela (1980). "An articulatory model for the vocal tracts of growing children". PhD thesis. Massachusetts Institute of Technology.

González, Julio (2004). "Formant frequencies and body size of speaker: a weak relationship in adult humans". In: *Journal of Phonetics* 32.2, pp. 277–287. ISSN: 0095-4470. DOI: https://doi.org/10.1016/S0095-4470(03)00049-4. URL: https://www.sciencedirect.com/science/article/pii/S0095447003000494.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, and Peter Trudgill (2004). *New Zealand English: its origins and evolution*. Cambridge University Press.

Gordon, Elizabeth and Margaret Maclagan (2001). "'Capturing a Sound Change': A Real Time Study Over 15 Years of the NEAR/SQUARE Diphthong Merger in New Zealand English". In: *Australian Journal of linguistics* 21.2, pp. 215–238.

Gordon, Elizabeth, Margaret Maclagan, and Jennifer Hay (2007). "The ONZE corpus". In: *Creating and digitizing language corpora*. Springer, pp. 82–104.

Gordon, Elizabeth and Peter Trudgill (1999). "Shades of things to come: Embryonic variants in New Zealand English sound changes". In: *English World-Wide* 20.1, pp. 111–124.

Gratton, Chantal (2016). "Resisting the gender binary: The use of (ING) in the construction of non-binary transgender identities". In: *University of Pennsylvania Working Papers in Linguistics* 22.2, p. 7.

Greisbach, Reinhold (2007). "Estimation of speaker height from formant frequencies". In: *International Journal of Speech Language and the Law* 6.2. Reprint. Originally published in 1999., pp. 265–277.

Griffin, Gabriele (2017). *social constructionism*. DOI: 10.1093/acref/9780191834837.013.0370. URL: https://www.oxfordreference.com/view/10.1093/acref/9780191834837.001.0001/acref-9780191834837-e-370.

Gubian, Michele, Jonathan Harrington, Mary Stevens, Florian Schiel, and Paul Warren (2019). "Tracking the New Zealand English NEAR/SQUARE merger using

functional principal components analysis". In: *Proceedings of the Annual Conference of the International Speech Communication Association*. Open Access Te Herenga Waka - Victoria University of Wellington, pp. 296–300.

Gurrieri, Lauren and Jenna Drenten (2021). "The feminist politics of choice: lipstick as a marketplace icon". In: *Consumption Markets & Culture* 24.3, pp. 225–240.

Hall, Kira (2013). "Commentary I: 'It'sa hijra!' Queer linguistics revisited". In: *Discourse & society* 24.5, pp. 634–642.

Halsey, Lewis G, Douglas Curran-Everett, Sarah L Vowler, and Gordon B Drummond (2015). "The fickle P value generates irreproducible results". In: *Nature methods* 12.3, pp. 179–185.

Hamley, Logan, Shiloh Groot, Jade Le Grice, Ashlea Gillon, Lara Greaves, Madhavi Manchi, and Terryann Clark (2021). ""You're the One That Was on Uncle's Wall!": Identity, Whanaungatanga and Connection for Takatāpui (LGBTQ+ Māori)". In: *Genealogy* 5.2, p. 54.

Hancock, Ange-Marie (2007). "When multiplication doesn't equal quick addition: Examining intersectionality as a research paradigm". In: *Perspectives on politics* 5.1, pp. 63–79.

Hashemi, Mohammad R and Raziye Eskandari (2017). "The learning of congruent and incongruent collocations utilizing dynamic assessment". In: *The Language Teacher* 41.6, pp. 9–14.

Hay, Jennifer, Katie Drager, and Paul Warren (2009). "Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English". In: *Australian Journal of Linguistics* 29.2, pp. 269–285.

Hay, Jennifer and Paul Foulkes (2016). "The evolution of medial/t/over real and remembered time". In: *Language* 92.2, pp. 298–330.

Hay, Jennifer, Janet Pierrehumbert, Abby Walker, and Patrick LaShell (2015). "Tracking word frequency effects through 130 years of sound change". In: *Cognition* 139, pp. 83–91.

Haynes, Erin F and Michael Taylor (2014). "An assessment of acoustic contrast between long and short vowels using convex hulls". In: *The Journal of the Acoustical Society of America* 136.2, pp. 883–891.

Hazenberg, Evan (2017). "Liminality as a lens on social meaning: A cross-variable analysis of gender in New Zealand English". PhD thesis. Victoria University of Wellington.

Henton, Caroline (1989). "Fact and fiction in the description of female and male pitch". In: *Language & Communication* 9.4, pp. 299–311. ISSN: 0271-5309. DOI: https://doi.org/10.1016/0271-5309(89)90026-8. URL: https://www.sciencedirect.com/science/article/pii/0271530989900268.

— (1992). "Acoustic variability in the vowels of female and male speakers." In: *The Journal of the Acoustical Society of America* 91.4, pp. 2387–2387.

— (1995). "Cross-language variation in the vowels of female and male speakers". In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol. 4. Acoustical Society of America, pp. 420–423.

Herman, Jody L (2013). "Gendered restrooms and minority stress: The public regulation of gender and its impact on transgender people's lives". In: *Journal of Public Management & Social Policy* 19.1, p. 65.

Hertlein, Katherine M, Erica E Hartwell, and Mashara E Munns (2016). "Attitudes toward bisexuality according to sexual orientation and gender". In: *Journal of Bisexuality* 16.3, pp. 339–360.

Hester, J David (2004). "Intersex (es) and informed consent: how physicians' rhetoric constrains choice". In: *Theoretical Medicine and Bioethics* 25.1, pp. 21–49.

Hillenbrand, James, Laura A Getty, Kimberlee Wheeler, and Michael J Clark (1994). "Acoustic characteristics of american english vowels". In: *The Journal of the Acoustical Society of America* 95.5, pp. 2875–2875.

Hlavac, Marek (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.2. Central European Labour Studies Institute (CELSI). Bratislava, Slovakia. URL: https://CRAN.R-project.org/package=stargazer.

Hogg, Margaret K and Jade Garrow (2003). "Gender, identity and the consumption of advertising". In: *Qualitative Market Research: An International Journal*.

Holland, Katie (Feb. 25, 2016). "Editorial: Toilet solution needed". In: *Rotorua Daily Post*. URL: https://www.nzherald.co.nz/rotorua-daily-post/opinion/news/article.cfm?c_id=1503435&objectid=11595533 (visited on 10/15/2019).

Hollien, Harry (1974). "On vocal registers". In: *Journal of Phonetics* 2.2, pp. 125–143.

Holmes, Janet (1987). "Hedging, fencing and other conversational gambits: an analysis of gender differences in New Zealand speech". In: *Women and language in Australian and New Zealand society*. Ed. by Anne Pawles, pp. 59–79.

— (1994). "New Zealand flappers: an analysis of T voicing in New Zealand English". In: *English World-Wide* 15.2, pp. 195–224.

— (1995). "Two for /t/: flapping and glottal stops in New Zealand English". In: *Te Reo*, pp. 53–72.

Holmes, Janet and Allan Bell (1992). "On shear markets and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English". In: *Language Variation and Change* 4.3, pp. 251–273.

Holmes, Janet and Stephanie Schnurr (2006). "'Doing femininity'at work: More than just relational practice 1". In: *Journal of Sociolinguistics* 10.1, pp. 31–51.

Honikman, Beatrice (1964). "Articulatory settings". In: *In honour of Daniel Jones: papers contributed on the occasion of his eightieth birthday, 12 September 1961*. Ed. by David Abercrombie, Dennis Butler Fry, Peter MacCarthy, Norman Carson Scott, and John Leslie Melville Trim. Longmans, pp. 73–84.

Hornsby, Michael (2008). "The incongruence of the Breton linguistic landscape for young speakers of Breton". In: *Journal of Multilingual and Multicultural Development* 29.2, pp. 127–138.

Human Rights Campaign (2013). *National Coming Out Day Youth Report*. [Online; accessed 9-November-2021]. URL: https://www.hrc.org/resources/youth-report.

— (2018). *2018 LGBTQ Youth Report*. [Online; accessed 9-November-2021]. URL: https://www.hrc.org/resources/2018-lgbtq-youth-report.

Hunter, Eric J, Kristine Tanner, and Marshall E Smith (2011). "Gender differences affecting vocal health of women in vocally demanding careers". In: *Logopedics Phoniatrics Vocology* 36.3, pp. 128–136.

International Olympic Committee (2021). *IOC Framework on fairness, inclusion and non-discrimination on the basis of gender identity and sex variations*. Document Release, Official Olympics Website. URL: https://stillmed.olympics.com/media/Documents/News/2021/11/IOC-Framework-Fairness-Inclusion-Non-discrimination-2021.pdf.

Ipsos (2021). *LGBT+ pride 2021 global survey: A 27-country Ipsos survey*. https://www.ipsos.com/sites/default/files/ct/news/documents/2021-06/lgbt-pride-2021-global-survey-ipsos.pdf.

Irvine, Judith T, Susan Gal, and Paul V Kroskrity (2009). "Language ideology and linguistic differentiation". In: *Linguistic anthropology: A reader* 1, pp. 402–434.

Isaacs, Dane and Leslie Swartz (2022). ""Stammering less so that I can be more of a man": Discourses of masculinities among young adult men in the Western Cape, South Africa, who stutter." In: *Psychology of Men & Masculinities* 23.1, p. 74.

Jacewicz, Ewa, Robert Allen Fox, and Joseph Salmons (2007). "Vowel space areas across dialects and gender". In: *16th International Congress of Phonetic Sciences, Saarbrücken, Germany*. Citeseer.

Jackson, Philip (1968). *Life in classrooms*. Teachers College Press.

Jackson, Sue and Tina Vares (2015). "New Visibilities: 'Using Video Diaries to Explore Girls' Experiences of Sexualized Culture". In: *Children, sexuality and sexualization*. Springer, pp. 307–320.

Jakobson, R., G. Fant, and M. Halle (1963). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*. Technical report. M.I.T. Press. URL: https://books.google.co.nz/books?id=dXVNAQAAMAAJ.

Jekel, James F (1977). "Should we stop using the P value in descriptive studies?" In: *Pediatrics* 60.1, pp. 124–126.

Jelenkovic, Aline, Reijo Sund, Yoon-Mi Hur, Yoshie Yokoyama, Jacob v B Hjelmborg, Sören Möller, Chika Honda, Patrik KE Magnusson, Nancy L Pedersen, Syuichi Ooki, et al. (2016). "Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts". In: *Scientific reports* 6.1, pp. 1–13.

Joel, Daphna, Ricardo Tarrasch, Zohar Berman, Maya Mukamel, and Effi Ziv (2014). "Queering gender: Studying gender identity in 'normative'individuals". In: *Psychology & Sexuality* 5.4, pp. 291–321.

Johnson, Amber and Robin M Boylorn (2015). "Digital media and the politics of intersectional queer hyper/in/visibility in Between Women". In: *Liminalities* 11.1, p. 1.

Johnson, Keith (1997). "Speech perception without speaker normalisation". In: *Talker variability in speech processing*, pp. 145–65.

— (2018a). "Speech production patterns in producing linguistic contrasts are partly determined by individual differences in anatomy". In: *UC Berkeley PhonLab Annual Report* 14.1.

— (2018b). "Vocal tract length normalization". In: *UC Berkeley PhonLab Annual Report* 14.1.

Johnson, Keith and Matthias J Sjerps (2021). "Speaker normalization in speech perception". In: *The handbook of speech perception*, pp. 145–176.

Jones, Angela (2020). "Where The Trans Men and Enbies At?: Cissexism, Sexual Threat, and the Study of Sex Work". In: *Sociology Compass* 14.2, e12750.

Jones, Jacq (2018). "Methodological Considerations for Incorporating Nonbinary Gender in Corpus Building". Presentation at the 2nd Workshop on Sociophonetic Variability on the English Varieties of Australia.

Keating, Patricia A, Marc Garellek, and Jody Kreiman (2015). "Acoustic properties of different kinds of creaky voice." In: *International Congress of Phonetic Sciences*. Vol. 2015. 1, pp. 2–7.

Kelley, Matthew C and Benjamin V Tucker (2020). "A comparison of four vowel overlap measures". In: *The Journal of the Acoustical Society of America* 147.1, pp. 137–145.

Kendrick, Karen (2008). "'Normalizing'female cancer patients: Look good, feel better and other image programs". In: *Disability & Society* 23.3, pp. 259–269.

Keshav, Aris (2021). "Stance, style, and intimacy: Interlocutor-based prosodic variation in a non-binary speaker". In: *Lavender Languages and Linguistics 27*.

Kim, Hyung-Tae (2017). "A new conceptual approach for voice feminization: 12 years of experience". In: *The Laryngoscope* 127.5, pp. 1102–1108.
King, Brian W (2019). *Communities of practice in language research: A critical introduction*. Routledge.
King, Helen (2016). *The one-sex body on trial: the classical and early modern evidence*. Routledge.
Kirby, Danielle (2016). "From pulp fiction to revealed text: a study of the role of the text in the Otherkin community". In: *Exploring religion and the sacred in a media age*. Routledge, pp. 155–168.
Kirk, Roger E (2001). "Promoting good statistical practices: Some suggestions". In: *Educational and Psychological Measurement* 61.2, pp. 213–218.
Kirkham, Sam (2015). "Intersectionality and the social meanings of variation: Class, ethnicity, and social practice". In: *Language in Society* 44.5, pp. 629–652.
Kirtley, M Joelle (2015). "Language, identity, and non-binary gender in Hawai'i". PhD thesis. University of Hawai'i at Manoa.
Knisely, Kris Aric (2020a). "Le français non-binaire: Linguistic forms used by non-binary speakers of French". In: *Foreign Language Annals* 53.4, pp. 850–876.
— (2020b). "Subverting the culturally unreadable: Understanding the self-positioning of non-binary speakers of French". In: *The French Review* 94.2, pp. 173–192.
Koziol, Michael (May 13, 2017). "Public service push for gender-neutral bathrooms". In: *The Sydney Morning Herald*. URL: https://www.smh.com.au/politics/federal/public-service-push-for-genderneutral-bathrooms-20170511-gw2taq.htm (visited on 10/15/2019).
Kuehn, Kathleen M and Sophie Parker (2021). "One of the blokes: Brewsters, branding and gender (in) visibility in New Zealand's craft beer industry". In: *Journal of Consumer Culture* 21.3, pp. 519–538.
Künzel, Hermann J (1989). "How well does average fundamental frequency correlate with speaker height and weight?" In: *Phonetica* 46.1-3, pp. 117–125.
Kwon, Ho-Beom (2010). "Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses". In: *The journal of advanced prosthodontics* 2.3, pp. 71–76.
Labov, William (1968). *A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Volume I: Phonological and Grammatical Analysis.* A report sponsored by the Office of Education (DHEW), Washington, DC. Bureau of Research.
— (1972). "Some principles of linguistic methodology". In: *Language in society* 1.1, pp. 97–120.
— (1973). "Sample questionnaire used by the project on linguistic change and variation". In: *University of Pennsylvania* 1.
— (1994). *Principles of Linguistic Change: Volume 1: Internal Factors*. Blackwell.
— (2006). *The social stratification of English in New York city*. Cambridge University Press.
Lacerda, Francisco (2003). "Phonology: An emergent consequence of memory constraints and sensory input". In: *Reading and Writing* 16.1, pp. 41–59.
Ladefoged, Peter and Keith Johnson (2014). *A course in phonetics*. Cengage learning.
Lammert, Adam C and Shrikanth S Narayanan (2015). "On short-time estimation of vocal tract length from formant frequencies". In: *PloS one* 10.7, e0132193.
Lamoureux, Mack (May 15, 2016). "Protest groups clash over Bill 10 at Alberta legislature grounds". In: *The Canadian Broadcasting Corporation*. URL: https://www.cbc.ca/news/canada/edmonton/bill-10-rallies-for-against-edmonton-1.3583002 (visited on 10/15/2019).

Langstrof, Christian (2006). "Acoustic evidence for a push-chain shift in the Intermediate Period of New Zealand English". In: *Language variation and change* 18.2, p. 141.

— (2011). "Vowel change as systemic optimisation: why the New Zealand English front vowel shift is not a good example". In: *English Language & Linguistics* 15.1, pp. 137–147.

Laqueur, Thomas (1992). *Making sex: Body and gender from the Greeks to Freud*. Harvard University Press.

Laver, John (1978). "The concept of articulatory settings: an historical survey". In: *Historiographia Linguistica* 5.1-2, pp. 1–14.

— (1980). "The phonetic description of voice quality". In: *Cambridge Studies in Linguistics London* 31, pp. 1–186.

Laver, John and Peter Trudgill (1979). *Phonetic and linguistic markers in speech. Social markers in speech, Ch. l, ed. by Klaus R. Scherer and Howard Giles, 1-32*.

Levon, Erez (2021). "Sexuality as non-binary: A variationist perspective". In: *The Routledge Handbook of Language, Gender, and Sexuality*, pp. 37–51.

Levon, Erez and Ronald Beline Mendes (2016). *Language, sexuality, and power: Studies in intersectional sociolinguistics*. Oxford University Press.

Lewis, Oscar (1941). "Manly-hearted women among the North Piegan". In: *American Anthropologist* 43.2, pp. 173–187.

Leys, Christophe, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata (2013). "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median". In: *Journal of experimental social psychology* 49.4, pp. 764–766.

"Fundamental Frequency, Pitch, F0" (2009). In: *Encyclopedia of Biometrics*. Ed. by Stan Z. Li and Anil Jain. Boston, MA: Springer US, pp. 592–592. ISBN: 978-0-387-73003-5. DOI: 10.1007/978-0-387-73003-5_775. URL: https://doi.org/10.1007/978-0-387-73003-5_775.

Lindeman, Meri (2021). "Situational variation in bigender Finnish speakers' idiolects". In: *Lavender Languages and Linguistics 27*.

Linhart, Pavel and Roman Fuchs (2015). "Song pitch indicates body size and correlates with males' response to playback in a songbird". In: *Animal Behaviour* 103, pp. 91–98.

Linke, CE (1973). "A study of pitch characteristics of female voices and their relationship to vocal effectiveness". In: *Folia Phoniatrica et Logopaedica* 25.3, pp. 173–185.

Livia, Anna (2005). "Fictional Representations of Butch/Femme Speech". In: *Gender articulated: Language and the socially constructed self* 245.

Loughlin, Ayden (2021). "Frequency of singular they for gender stereotypes and the influence of the queer community". In: *Lavender Languages and Linguistics 27*.

Maclagan, Margaret and Jennifer Hay (2004). "The rise and rise of New Zealand English DRESS". In: *The Proceedings of the Australian International Conference on Speech Science and Technology*, pp. 183–188.

— (2007). "Getting fed up with our feet: Contrast maintenance and the New Zealand English" short" front vowel shift". In: *Language variation and change* 19.1, p. 1.

Maclagan, Margaret A and Elizabeth Gordon (1996). "Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger". In: *Language Variation and Change* 8.1, pp. 125–147.

— (2004). "The story of New Zealand English: What the ONZE project tells us". In: *Australian Journal of Linguistics* 24.1, pp. 41–56.

Maclagan, Margaret A, Elizabeth Gordon, and Gillian Lewis (1999). "Women and sound change: Conservative and innovative behavior by the same speakers". In: *Language variation and change* 11.1, pp. 19–41.

Mair, Ron (Nov. 2, 2021). Personal Communication with Statistics New Zealand.

Martin, Karin A (1998). "Becoming a gendered body: Practices of preschools". In: *American sociological review*, pp. 494–511.

McConnell-Ginet, Sally (1978). "Intonation in a man's world". In: *Signs: Journal of Women in Culture and Society* 3.3, pp. 541–559.

McDonald, Lisa (Aug. 11, 2017). "Gender issues left out of New Zealand's next census?" In: *Stuff*. URL: https://www.stuff.co.nz/national/79760678/are-new-zealand-bathrooms-doing-enough-to-accommodate-transgender-people (visited on 09/05/2017).

McElligott, AG, M Birrer, and E Vannoni (2006). "Retraction of the mobile descended larynx during groaning enables fallow bucks (Dama dama) to lower their formant frequencies". In: *Journal of zoology* 270.2, pp. 340–345.

McSweeny, Philip (May 12, 2016). "Are New Zealand bathrooms doing enough to accommodate Transgender people?" In: *Stuff*. URL: https://www.stuff.co.nz/national/79760678/are-new-zealand-bathrooms-doing-enough-to-accommodate-transgender-people (visited on 09/16/2019).

Micancin, Jonathan P and R Haven Wiley (2014). "Allometric convergence, acoustic character displacement, and species recognition in the syntopic cricket frogs Acris crepitans and A. gryllus". In: *Evolutionary Biology* 41.3, pp. 425–438.

Milroy, Lesley (1980). "Language and social networks". In.

Milroy, Lesley and Carmen Llamas (2013). "Social networks". In: *The handbook of language variation and change*, pp. 407–427.

Minto, Catherine L, Lih-Mei Liao, Christopher RJ Woodhouse, Phillip G Ransley, and Sarah M Creighton (2003). "The effect of clitoral surgery on sexual outcome in individuals who have intersex conditions with ambiguous genitalia: a cross-sectional study". In: *The Lancet* 361.9365, pp. 1252–1257.

Mitchell, Mani B (2018). *2018 LGBTQ Youth Report*. [Transcript of a speech given at the beehive banquet room in Wellington to commemorate the 30th anniversary of homosexual law reform in New Zealand. Online; accessed 9-November-2021]. URL: http://www.ianz.org.nz/2016/04/30th-anniversary-homosexual-law-reform/.

Mondorf, Britta (2002). "Gender differences in English syntax". In: *Journal of English Linguistics* 30.2, pp. 158–180.

Munson, Benjamin (2007). "Lexical characteristics mediate the influence of sex and sex typicality on vowel-space size". In: *Proceedings of the 16th International Congress of Phonetic Sciences*. Citeseer, pp. 885–888.

Munson, Benjamin and Molly Babel (2007). "Loose lips and silver tongues, or, projecting sexual orientation through speech". In: *Language and Linguistics Compass* 1.5, pp. 416–449.

Murry, Thomas and Sadanand Singh (1980). "Multidimensional analysis of male and female voices". In: *The journal of the Acoustical society of America* 68.5, pp. 1294–1300.

"Mutual Exclusivity Assumption" (2022). In: *The American Psychological Association Dictionary*. The American Psychological Association. URL: https://dictionary.apa.org/mutual-exclusivity-assumption.

Myers, James (1995). "The categorical and gradient phonology of variable t-deletion in English". In: *Paper presented at the International Workshop on Language Variation and Linguistic Theory*. Nijmegen, Netherlands.

Myketiak, Chrystie (2016). "Fragile masculinity: Social inequalities in the narrative frame and discursive construction of a mass shooter's autobiography/manifesto". In: *Contemporary Social Science* 11.4, pp. 289–303.

Neel, Amy (2008). "Vowel Space Characteristics and Vowel Identification Accuracy". In: *Journal of Speech, Language, and Hearing Research* 51.3, pp. 574–585.

Nelson, Stephen A (2011). *2 Component Phase Diagrams (EENS 2110 Minerology Tutorial)*. URL: https://www.tulane.edu/~sanelson/eens211/2compphasdiag.html.

Ng, Manwa L, Gigi Hsueh, and Cheung-Shing Sam Leung (2010). "Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children". In: *International journal of speech-language pathology* 12.3, pp. 230–236.

Oates, Jenni and Giorgia Dacakis (1997). "Voice change in transsexuals". In: *Venereology* 10.3.

Ochs, Eleanor (1992). "Indexing gender". In: *Rethinking context: Language as an interactive phenomenon. Edited by Alessandro Duranti & Charles Goodwin* 11, pp. 335–358.

Papadopoulos, Ben (2021). "Queer Speakers and Gendered Language: A New Linguistic Gender Typology". In: *Lavender Languages and Linguistics* 27.

Papp, Viktoria (2011). "The female-to-male transsexual voice: Physiology vs. performance in production". PhD thesis. Rice University.

Parish-Morris, Julia, Mark Y Liberman, Christopher Cieri, John D Herrington, Benjamin E Yerys, Leila Bateman, Joseph Donaher, Emily Ferguson, Juhi Pandey, and Robert T Schultz (2017). "Linguistic camouflage in girls with autism spectrum disorder". In: *Molecular autism* 8.1, pp. 1–12.

Pascoe, Cheri J (2011). *Dude, you're a fag: Masculinity and sexuality in high school*. University of California Press.

Pearce, Jo (2019). "Identity, socialization and environment in transgender speakers: Sociophonetic variation in creak and/s". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, pp. 3290–3294.

Pebesma, Edzer (2018). "Simple Features for R: Standardized Support for Spatial Vector Data". In: *The R Journal* 10.1, pp. 439–446. DOI: 10.32614/RJ-2018-009. URL: https://doi.org/10.32614/RJ-2018-009.

Peirce, Charles Sanders (1977). *Semiotics and significs, ed. Charles Hardwick*.

Pépiot, Erwan (2009). "The making of French vocalic triangles: the case of a woman's voice versus a man's voice". Mémoire de recherche de Master 1 (50 p.). Sous la direction du Professeur Jean-Yves Dommergues. MA thesis. Université Paris 8.

— (2015). "Voice, speech and gender:. male-female acoustic differences and cross-language variation in english and french speakers". In: *Corela. Cognition, représentation, langage* HS-16.

Peters, Jeremy W., Jo Becker, and Julie Hirschfeld Davis (Feb. 22, 2017). "Trump Rescinds Rules on Bathrooms for Transgender Students". In: *The New York Times*. URL: https://www.nytimes.com/2017/02/22/us/politics/devos-sessions-transgender-students-rights.html (visited on 11/22/2019).

Peterson, Gordon E (1951). "The phonetic value of vowels". In: *Language*, pp. 541–553.

Peterson, Gordon E and Harold L Barney (1952). "Control methods used in a study of the vowels". In: *The Journal of the acoustical society of America* 24.2, pp. 175–184.

Petre, Caitlin, Brooke Erin Duffy, and Emily Hund (2019). ""Gaming the system": Platform paternalism and the politics of algorithmic visibility". In: *Social Media+ Society* 5.4, p. 2056305119879995.

Pettinato, Michèle, Outi Tuomainen, Sonia Granlund, and Valerie Hazan (2016). "Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication". In: *Journal of Phonetics* 54, pp. 1–14. ISSN: 0095-4470.

DOI: https://doi.org/10.1016/j.wocn.2015.07.002. URL: https://www.sciencedirect.com/science/article/pii/S0095447015000522.

Pisanski, Katarzyna, Benedict C Jones, Bernhard Fink, Jillian JM O'Connor, Lisa M DeBruine, Susanne Röder, and David R Feinberg (2016). "Voice parameters predict sex-specific body morphology in men and women". In: *Animal behaviour* 112, pp. 13–22.

Plavcan, J Michael (2012). "Body size, size variation, and sexual size dimorphism in early Homo". In: *Current Anthropology* 53.S6, S409–S423.

Podesva, Robert J (2007). "Phonation type as a stylistic variable: The use of falsetto in constructing a persona". In: *Journal of sociolinguistics* 11.4, pp. 478–504.

— (2008). "Three sources of stylistic meaning". In: *Texas Linguistic Forum*. Vol. 51, pp. 1–10.

Podesva, Robert J and Patrick Callier (2015). "Voice quality and identity". In: *Annual review of applied Linguistics* 35, pp. 173–194.

Poncet, Paul (2019). *modeest: Mode Estimation*. R package version 2.4.0. URL: https://CRAN.R-project.org/package=modeest.

Portela, Annika Szabo, Britta Hammarberg, and Maria Södersten (2013). "Speaking fundamental frequency and phonation time during work and leisure time in vocally healthy preschool teachers measured with a voice accumulator". In: *Folia Phoniatrica et Logopaedica* 65.2, pp. 84–90.

Pratt, Teresa (2020). "Embodying "tech": Articulatory setting, phonetic variation, and social meaning". In: *Journal of Sociolinguistics* 24.3, pp. 328–349.

Pratt, Teresa and Annette D'Onofrio (2017). "Jaw setting and the California Vowel Shift in parodic performance". In: *Language in Society* 46.3, pp. 283–312.

Przewozny, Anne and Cécile Viollain (2016). "On the representation and evolution of Australian English and New Zealand English". In: *Anglophonia. French Journal of English Linguistics* 21.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ramet, Sabrina Petra, ed. (1996). *Gender reversals and gender cultures: Anthropological and historical perspectives*. Psychology Press.

Recasens, Daniel (2010). "Differences in base of articulation for consonants among Catalan dialects". In: *Phonetica* 67.4, pp. 201–218.

Rechsteiner, Jack (2021). "Nonbinary speakers' use of (ING) across gender-related topics". In: *NWAV44 Texas*.

Rendall, Drew, Sophie Kollias, Christina Ney, and Peter Lloyd (2005). "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry". In: *The Journal of the Acoustical Society of America* 117.2, pp. 944–955.

Rippon, Gina, Rebecca Jordan-Young, Anelis Kaiser, and Cordelia Fine (2014). "Recommendations for sex/gender neuroimaging research: key principles and implications for research design, analysis, and interpretation". In: *Frontiers in human neuroscience* 8, p. 650.

Ruff, Christopher (2002). "Variation in human body size & shape". In: *Annual review of anthropology* 31.1, pp. 211–232.

Ryalls, John H and Philip Lieberman (1982). "Fundamental frequency and vowel perception". In: *The Journal of the Acoustical Society of America* 72.5, pp. 1631–1634.

Salih, Sara (2007). "On Judith butler and performativity". In: *Sexualities and communication in everyday life: A reader*, pp. 55–68.

Sa'ar, Amalia (2007). "Masculine talk: On the subconscious use of masculine linguistic forms among Hebrew-and Arabic-speaking women in Israel". In: *Signs: Journal of Women in Culture and Society* 32.2, pp. 405–429.

Schmidt, Johanna (2001). "Redefining fa'afafine: Western discourses and the construction of transgenderism in Samoa". In: *Intersections: Gender, History and Culture in the Asian Context* 6, pp. 19–31.

Schroeder, Manfred R (2013). *Computer speech: recognition, compression, synthesis*. Vol. 35. Springer Science & Business Media.

Secombe, Wally (1974). "The Housewife and her Labour under Capitalism". In: *New Left Review* 83, p. 3.

Setchell, Joanna M, Phyllis C Lee, E Jean Wickings, and Alan F Dixson (2001). "Growth and ontogeny of sexual size dimorphism in the mandrill (Mandrillus sphinx)". In: *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 115.4, pp. 349–360.

Silverstein, Michael (1976). "Shifters, linguistic categories, and cultural description". In: *Meaning in Anthropology*, pp. 11–55.

— (2003). "Indexical order and the dialectics of sociolinguistic life". In: *Language & communication* 23.3-4, pp. 193–229.

Simpson, Adrian (2011). "Is there a relationship between acoustic vowel space size and fundamental frequency?" In: *International Congress of Phonetic Sciences*, pp. 1854–1857.

Simpson, Adrian and Christine Ericsdotter (2007). "Sex-specific differences in f0 and vowel space". In: *XVIth international congress of phonetic sciences*, pp. 933–936.

Singh, Sadanand and Thomas Murry (1978). "Multidimensional classification of normal voice qualities". In: *The Journal of the Acoustical Society of America* 64.1, pp. 81–87.

Skuk, V. G. and S. R. Schweinberger (2014). "Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender". In: *Journal of Speech, Language, and Hearing Research* 57.1, pp. 285–296.

Smaldino, Paul E and Matthew A Turner (2020). "Covert signaling is an adaptive communication strategy in diverse populations". In: Preprint. DOI: 10.1037/rev0000344. URL: osf.io/preprints/socarxiv/j9wyn.

Smith, Eliot R and Michael A Zarate (1992). "Exemplar-based model of social judgment." In: *Psychological review* 99.1, p. 3.

Smith, Phill (Aug. 12, 2021). "Births, deaths, and genders: a quiet bill goes loud". In: *Radio New Zealand*. URL: https://www.rnz.co.nz/national/programmes/the-house/audio/2018807955/births-deaths-and-genders-a-quiet-bill-goes-loud (visited on 11/20/2021).

Statistics New Zealand (2015). *Gender identity standard world-first in New Zealand*. [Online; accessed 5-September-2017]. URL: http://www.stats.govt.nz/tools_and_services/media-centre/media-releases-2015/gender-identity-17-july-15.aspx.

— (2016). *Statistical Standard for Sex*. [Online; accessed 5-September-2017]. URL: http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/sex.aspx.

Steele, Ariana J (2019). "Non-binary speech, race, and non-normative gender: sociolinguistic style beyond the binary". MA thesis. The Ohio State University.

Stevens, Kenneth N, Stanley Kasowski, and C Gunnar M Fant (1953). "An electrical analog of the vocal tract". In: *The Journal of the Acoustical Society of America* 25.4, pp. 734–742.

Storoshenko, Dennis Ryan (2021). "Online Communities of Use for Neoreflexives". In: *Lavender Languages and Linguistics 27*.

Strand, Elizabeth A (1999). "Uncovering the role of gender stereotypes in speech perception". In: *Journal of language and social psychology* 18.1, pp. 86–100.

Strömbergsson, Sofia (2016). "Today's Most Frequently Used F0 Estimation Methods, and Their Accuracy in Estimating Male and Female Pitch in Clean Speech." In: *Interspeech*. Dresden, pp. 525–529.

Stuart-Smith, Jane (1999). "Voice quality in Glaswegian". In: *Proceedings of the International Congress of Phonetic Sciences*. Vol. 14, pp. 2553–2556.

Stuart-Smith, Jane, Claire Timmins, Farhana Alam, Frans Gregersen, Jeffrey Parrot, and Pia Quist (2011). "Hybridity and ethnic accents". In: *Selected Papers ICLaVE Amsterdam: Benjamins* 5, pp. 43–57.

Styler, Will (2013). "Using Praat for linguistic research". In: *University of Colorado at Boulder Phonetics Lab*.

Tagliamonte, Sali A (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Takefuta, Yukio, Elizabeth G Jancosek, and Michael Brunt (1972). "A statistical analysis of melody curves in the intonation of American English". In: *Proceedings of the seventh International Congress of Phonetic Sciences/Actes du Septième Congrès international des sciences phonétiques*. De Gruyter Mouton, pp. 1035–1039.

Talkin, David (2015). "REAPER: Robust epoch and pitch estimator". In: *Github: https://github. com/google/REAPER*.

Tarrant, Shira (2013). *When sex became gender*. Routledge.

Taylor, Ben (1996). *Gay men, femininity and/t/in New Zealand English*. Tech. rep. Wellington Working Papers in Linguistics.

The Language Archive (2021). *ELAN (EUDICO Linguistic Annotator)*. Computer Software. Version 6.2. Max Planck Institute for Psycholinguistics, The Language Archive. Nijmegen, The Netherlands. URL: https://archive.mpi.nl/tla/elan.

Titman, Nat (2014). *How many people in the United Kingdom are nonbinary?* [Online; accessed 9-November-2021]. URL: https://practicalandrogyny.com/2014/12/16/how-many-people-in-the-uk-are-nonbinary/.

Todd, Simon John (2019). "The listener in language change". PhD thesis. Stanford University.

Tollfree, Laura (2001). "Variation and change in Australian English consonants". In: *Blair & Collins (eds.)*, pp. 45–67.

Tramo, Mark Jude, Gaurav D Shah, and Louis D Braida (2002). "Functional role of auditory cortex in frequency processing and pitch perception". In: *Journal of neurophysiology* 87.1, pp. 122–139.

Trudgill, Peter (1972). "Sex, covert prestige and linguistic change in the urban British English of Norwich". In: *Language in society* 1.2, pp. 179–195.

— (2004). *New-dialect formation: The inevitability of colonial Englishes*. Oxford University Press, USA.

Tsao, Ying-Chiao, Gary Weismer, and Kamran Iqbal (2006). "The effect of intertalker speech rate variation on acoustic vowel space". In: *The Journal of the Acoustical Society of America* 119.2, pp. 1074–1082.

T'Sjoen, Guy, Mieke Moerman, John Van Borsel, Els Feyen, Robert Rubens, Stanislas Monstrey, Piet Hoebeke, Petra De Sutter, and Griet De Cuypere (2006). "Impact of voice in transsexuals". In: *International Journal of Transgenderism* 9.1, pp. 1–7.

UMR Research Ltd. (2016). *Public Attitudes and Understandings of Ethnic Identity – A Qualitative Study*. [Online; accessed 5-September-2017]. URL: http://www.stats.

govt.nz/browse_for_stats/people_and_communities/Households/umr-research-study.aspx.

Valderrama, Joaquin T, Elizabeth F Beach, Mridula Sharma, Shivali Appaiah-Konganda, and Elaine Schmidt (2021). "Design and evaluation of the effectiveness of a corpus of congruent and incongruent English sentences for the study of event related potentials". In: *International Journal of Audiology* 60.2, pp. 96–103.

Van Borsel, John, Griet De Cuypere, and Hilde Van den Berghe (2001). "Physical appearance and voice in male-to-female transsexuals". In: *Journal of Voice* 15.4, pp. 570–575.

Van Borsel, John, Kathelijne de Pot, and Griet De Cuypere (2009). "Voice and physical appearance in female-to-male transsexuals". In: *Journal of Voice* 23.4, pp. 494–497.

Van Dommelen, Wim A and Bente H Moxness (1995). "Acoustic parameters in speaker height and weight identification: sex-specific behaviour". In: *Language and speech* 38.3, pp. 267–287.

Veale, Jaimie, Jack Byrne, Kyle KH Tan, Sam Guy, Ashe Yee, Tāwhanga Mary-Legs Nopera, and Ryan Bentham (2019). *Counting Ourselves: The health and wellbeing of trans and non-binary people in Aotearoa New Zealand*. Transgender Health Research Lab.

Villarreal, Dan, Lynn Clark, Jennifer Hay, and Kevin Watson (2020). "From categories to gradience: Auto-coding sociophonetic variation with random forests". In: *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11.1, p. 6.

Vogel, Adam P, Paul Maruff, Peter J Snyder, and James C Mundt (2009). "Standardization of pitch-range settings in voice acoustic analysis". In: *Behavior research methods* 41.2, pp. 318–324.

Vorperian, Houri K and Ray D Kent (2007). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data". In: *Journal of Speech, Language, and Hearing Research*.

Walker, Abby and Jennifer Hay (2011). "Congruence between 'word age'and 'voice age'facilitates lexical access". In.

Walsh, Bridget and Anne Smith (2002). "Articulatory movements in adolescents: evidence for protracted development of speech motor control processes". In: *Journal of Speech, Language, and Hearing Research* 45.6, pp. 1119–1133.

Ward, Colleen and Anne-Marie Masgoret (2008). "Attitudes toward immigrants, immigration, and multiculturalism in New Zealand: A social psychological analysis". In: *International Migration Review* 42.1, pp. 227–248.

Weirich, Melanie and Adrian Simpson (2013). "Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size". In: *The Journal of the Acoustical Society of America* 134.4, pp. 2965–2974.

White, Hannah, Joshua Penney, Andy Gibson, Anita Szakay, and Felicity Cox (2022). "Evaluating automatic creaky voice detection methods". In: *The Journal of the Acoustical Society of America* 152.3, pp. 1476–1486.

Whiteside, Sandra P (1998). "Identification of a speaker's sex: a study of vowels". In: *Perceptual and Motor Skills* 86.2, pp. 579–584.

— (2001). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study". In: *The Journal of the Acoustical Society of America* 110.1, pp. 464–478.

"Whole Object Assumption" (2022). In: *The American Psychological Association Dictionary*. The American Psychological Association. URL: https://dictionary.apa.org/whole-object-assumption.

Wieling, Martijn and Mark Tiede (2017). "Quantitative identification of dialect-specific articulatory settings". In: *The Journal of the Acoustical Society of America* 142.1, pp. 389–394.

Wilson, Ian and Bryan Gick (2014). "Bilinguals use language-specific articulatory settings". In: *Journal of Speech, Language, and Hearing Research* 57.2, pp. 361–373.

Winkelmann, Raphael, Klaus Jaensch, Steve Cassidy, and Jonathan Harrington (2021). *emuR: Main Package of the EMU Speech Database Management System*. R package version 2.2.0.

Winkler, Ralf, Susanne Fuchs, and Pascal Perrier (2006). "The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modelling". In: *Proceedings of the International Seminar on Speech Production*. CEFALA Ubatuba, Brazil, pp. 509–516.

Witt, Susan D (1997). "Parental influence on children's socialization to gender roles". In: *Adolescence* 32.126, p. 253.

Wolfe, Virginia I, David L Ratusnik, Furman H Smith, and Gretajo Northrop (1990). "Intonation and fundamental frequency in male-to-female transsexuals". In: *Journal of Speech and Hearing Disorders* 55.1, pp. 43–50.

Woods, Nicola J (1997). "The formation and development of New Zealand English: Interaction of gender-related variation and linguistic change". In: *Journal of Sociolinguistics* 1.1, pp. 95–125.

Yang, Byunggon (1990). "Development of vowel normalization procedures: English and Korean". PhD thesis. University of Texas at Austin.

Yorke, Harry (Mar. 6, 2017). "No more separate boys' and girls' toilets at primary schools in Glasgow to help pupils who are 'confused about their gender'". In: *The Telegraph*. URL: https://www.telegraph.co.uk/education/2017/03/06/primary-schools-will-have-unisex-bathrooms-council-reveals/ (visited on 10/15/2019).

Zhang, Zhaoyan (2016). "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model". In: *The Journal of the Acoustical Society of America* 139.4, pp. 1493–1507.

Zimman, Lal (2013). "Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men". In: *Journal of Language and Sexuality* 2.1, pp. 1–39.

— (2014). "The discursive construction of sex". In: *Queer excursions: Retheorizing binaries in language, gender, and sexuality*, pp. 13–34.

— (2017). "Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/". In: *Language in Society* 46.3, pp. 339–370.

— (2021). "Gender diversity and the voice". In: *The Routledge Handbook of Language, Gender, and Sexuality*. Routledge, pp. 69–90.

Zimman, Lal and Will Hayworth (2020). "How we got here: Short-scale change in identity labels for trans, cis, and non-binary people in the 2000s". In: *Proceedings of the Linguistic Society of America* 5.1, pp. 499–513.

— (2021). "You just got transgendered!': Lexical change and language ideologies in discourse about identity labels for trans, cis and non-binary people". In: *Lavender Languages and Linguistics 27*.

Zwicker, Eberhard (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)". In: *The Journal of the Acoustical Society of America* 33.2, pp. 248–248.