Can women "sound gay"?
A sociophonetic study of /s/ and pitch of gay and straight
British-English speaking women

Salina Cuddy

PhD

University of York
Language and Linguistic Science

December 2019

Abstract

This thesis presents a third wave variationist project, influenced by queer linguistics, that considers the idea of a "gay voice" for female English speakers in Yorkshire, England. This work demonstrates the complexity of identity and why it is important that researchers look past macro-social categories alone, and consider the nuance of local communities.

Two studies were conducted to investigate the relationship between women's sexual orientation and speech: a production study and a perception study. The production study analysed data from 22 participants, from Yorkshire, who identified as female, white, between the ages of 21 and 46, and 12 identified as gay and 10 as straight. It was found that while there were some significant differences between gay and straight speakers based on F0 and /s/ centre of gravity measurements, there were more substantial differences between participants that were both gay and part of a local football team than participants that were either straight or not on the team.

The second study on the perception of a "gay voice" for women asked participants to rate qualities such as "homosexual" and "feminine" in order to understand how stimuli with digitally altered F0 would be perceived. Data reveals that sentences with a lowered F0 were perceived as sounding more "homosexual" than stimuli with a raised F0, indicating that a lower F0 may be associated with a "gay voice" for women. However, the nuance of the participants had a significant impact on how they rated these characteristics.

This in-depth analysis of how speakers may perform their sexual orientation and how listeners may perceive sexual orientation demonstrates the complexity of language and identity. It was found that local communities are significant in presenting a specific gay identity and how important it is to look at the intersection of many factors to fully understand linguistic variation.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank the extensive network of support I received during my PhD process, because without that support this would not be possible. To my supervisors, Paul Kerswill and Dom Watt, thank you for your guidance and input in the development of this work. You were both there from the creation of this idea and your enthusiasm for the project and belief in my abilities was vital in getting to this point.

I would also like to thank Lucy Jones for all of her wonderful guidance and feedback. You went above and beyond in helping me understand the theoretical concepts and developing my work. I sincerely appreciate your time and effort.

Thank you to the Language and Linguistic Science Department who were there for me whenever I needed help. There is so much knowledge in the department and I am fortunate that I was able to call on it frequently. In particular, thank you to Grace Wood for helping me work through ideas and reading my drafts – you made all the difference. And thank you to Justin Lo for all the amazing technological help.

I want to extend a massive thank you to my network in the Berrick Saul Building. Rosemary Alexander-Jones, Marie Allitt, Amy Creighton, Jennie England, Max Gee, and Andrew Stead: you were endlessly supportive of me. You are all brilliant and I am lucky to have you in my life.

To my family away from home: Meaghan Mackie, Alice Taylor, Danielle Hyde, and Ashley Timm, you all made this possible by giving me more love and support than I could have ever expected. To my family and especially my mom and dad, thank you for being with me on this wild journey. What I was jabbering on about might not have always made much sense, but you listened and provided constant support and love, even from a distance. I am the person I am because of you and I am so grateful. Finally, to Simon Quinn, thank you for being my rock throughout all of this. You stuck with me through the ups and downs, and I am truly blessed to have you by my side.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1. Introduction

This research began by seeking to understand the influence sexual orientation has on speech production among female British English speakers based in Northern England. In the course of completing this work, it became clear that it is also essential to demonstrate the importance of locally salient communities and to show the benefits of a third wave variationist approach to sociolinguistic research, à la Eckert (2012). In order to achieve both of these goals, this thesis presents a two-part study that focuses on two phonetic features produced by both gay and straight speakers from Yorkshire and then a discussion on the importance of the methodology and analysis that was employed in these studies.

Research on language and sexual orientation is inherently complex, as it lies at the intersection of multiple disciplines, and these interactions can be highly variable between individual speakers. It is also important when researching sexual orientation to remember that sexual orientation is not the only significant facet of identity to individuals, and therefore there are also intersections between sexual orientation, gender, social class, nationality, and a myriad of other factors. I will attempt to place this particular group of speakers in a clear context and discuss how these individuals may use certain phonetic features to index sexual orientation. This project was heavily influenced by intersectionality theory and the discussions of the individual studies, as well as the thesis as a whole, indicate why it is critical to consider the intersecting identities participants have.

The present study is interested in the idea of a "gay voice" for women. In the context of this thesis, a gay voice refers to a set of features that may index a gay sexual orientation for female speakers. It is important to clarify that not every gay speaker would inherently have a gay voice, or that every person that has a gay voice would be gay. Instead it refers to the social perception that listeners may have when hearing this particular set of features from a female speaker. The present study is interested in exploring some of the features that may be part of a gay voice set.

While there are many features that could constitute a gay voice, this thesis focuses on two phonetic variables: F0 and /s/. These features were chosen based on previous language and gender studies, as well as language and sexuality studies. As will be developed in section 2.5, F0 and /s/ have both regularly been considered when researching gender differences in sociolinguistic research and these associations with gender have led to them being commonly studied in relation to sexual orientation.

This project was influenced by previous third wave variationist research (including: Eckert, 2000; Levon, 2006; Podesva, 2007; Moore and Podesva, 2009; Holmes-Elliott and Levon, 2017). In order to consider the complexity not only of sexual orientation, but identity more generally, concepts such as stance, indexicality, and performativity were drawn on throughout the research. These theories of identity construction also demonstrate that identity is in no way fixed, and instead is created in the moment. This is also the belief in queer linguistics, which was another influential theoretical framework. Finally, this research was strongly influenced by intersectionality theory, as described by Levon (2015) in a specific sociolinguistic context.

Though sexual orientation was the original interest in the research, it was discovered that a locally significant group surrounding a women's football team was more influential than sexual orientation alone. I argue throughout this entire thesis that in order to fully develop our understanding of linguistic variation and the interactions between identity and language, sociolinguists must include more third wave approaches that consider the intersection of multiple identities. These intersections exist and are important not just in language, gender, and sexuality research, but any research that considers human participants. Further, these intersections are equally important to consider for all groups and not only those that consist of typically minority identities (such as gay women or Black, Asian, and minority ethnic communities within the UK). While the specific phonetic findings will not be expanded to larger scale groups (such as a pan-English gay voice for women), the methodology and analysis can be considered for research outside gender and sexuality linguistic work.

1.1 Constructs and Terminology

Before considering the research, it is important to understand the terminology that is used throughout the thesis. There is an elaborate lexicon that attempts to be both inclusive and descriptive when discussing sexuality, but it can be complicated if terms are not defined or if they are used interchangeably.

As Cameron and Kulick describe in their book *Language and Sexuality* (2003), the term *sexuality* is fluid and may refer to the gender one is attracted to, but also the fantasies one may have and other elements of sexual attraction. Therefore, *sexual orientation* will be used to specifically describe the attraction a person has to a certain gender. When discussing sexual orientations, this study uses the term *straight* to describe heterosexual people and *gay* when describing homosexual people that are both male and female. The choice to use *gay* when discussing women is based on the data collected from the production study, and this will be expanded on in Section 3.3.1.1. *Sexuality* will refer to sexual and romantic attraction in a more general sense, the way Cameron and Kulick describe.

While this thesis focuses primarily on gay speakers, there are references to other gender identities and sexual orientations. When referring to the larger community of people that do not identify as straight or who have a different gender than the one assigned at birth, this thesis uses the acronym *LGBTQ* (lesbian, gay, bisexual, transgender, and queer). I use it as a way to include all non-normative and/or marginalised genders, romantic and sexual identities, and intersex people. As will be seen in the discussion on the word *queer* in Section 2.3.2, it can prove difficult to settle on a term that is inclusive of all people, and therefore, I will be using LGBTQ broadly.

When discussing sexual orientation, we also need to clarify the use of *sex* and *gender* both in this thesis and how it has been used in previous research. As set forth by Lips (2005, p. 6) in her book *Sex & Gender: An Introduction, sex* typically refers to "[the] anatomy and the classification of individuals based on their anatomical category". *Gender* generally refers to the identity someone associates

with and presents, which may not always align with the sex one was born with. Eckert & McConnell-Ginet write (1992, p. 471), "Gender can be thought of as a sex-based way of experiencing other social attributes such as class, ethnicity, or age (and also less obviously social qualities like ambition, athleticism, musicality, and the like)".

However, as Butler (1990, pp. 10–11) writes, it is not this clean cut. She questions where the physical view of sex begins, if it is based on anatomy, or chromosome differences, or even hormonal differences. This level of biological scrutiny is even more complicated, as many biologists agree that sex is more of a spectrum than a binary categorisation (Ainsworth, 2015). It is possible, for example, for a person to have XX chromosomes, typically associated with female sex characterises, but to develop testes. This discussion is to highlight the complexity of sex and gender and to reinforce the importance of the gender a person performs.

Due to the complex nature of sex and gender, the present thesis will typically refer to a participant's gender exclusively. The prefix *cis-* or sometimes the word *cis-gender* will also be used throughout this thesis, and this refers to a person who identifies with the same gender that they were assigned at birth.

Gender must also be kept distinct from the concepts of masculinity and femininity. *Masculinity* refers to attributes and behaviours commonly associated with men, which can include independence, competitiveness, and aggression. F*emininity* refers to attributes and behaviours commonly associated with women, which include passivity, sensitivity, and gentleness. As it is possible for both masculine and feminine qualities to be presented or perceived by any gender, they must be thought of separately from one's gender.

Though this list of terminology is extensive, it is important to have a clear understanding of how these terms will be used throughout the thesis in order to ensure there is no confusion.

4

1.2 Outline of the Thesis

The research will be presented as follows. First, in Chapter 2, there will be a review of the previous work that is relevant to the present thesis. To begin, there will be a discussion on the different approaches to variationist sociolinguistics, as well as some relevant sociolinguistic theory that is used throughout this thesis. After that, there will be a more specific view of research on language and sexuality, as well as a discussion of queer theory and queer linguistics. Following this presentation, there will be a review on some of the important phonetic contexts this thesis works in, namely Yorkshire accents, F0, and fricatives. Finally, there will be a focused review of the previous research that includes gay women and how this thesis seeks to fill gaps that are currently present.

Chapter 3 will present the first study of this thesis, which is a study of speech production patterns among a group of Yorkshire speakers. This initial study considers the speech of members of a local football team, as well as non-team members from similar geographic areas. This chapter will include the data collection methodology, the results, and a discussion of the insights of this particular study. This chapter shows that while there are features that are significantly different between the gay and straight participants, one must consider local communities as these appear to be more important.

Chapter 4 will present the perception study that follows from the production study. The perception study was created based on insights from the production study and seeks to investigate if listeners perceive a gay voice. This chapter will include the methodology for collecting the stimuli and the creation of the survey, as well as a presentation of the results and a discussion of their implications. As with Chapter 3, this perception study exemplifies the importance of a nuanced view of the participants, as there were significantly different results based on the identities of the participants of the survey.

Chapter 5 will consider both the production and perception studies together and draw some conclusions based on patterns that appear across the studies. Some of these considerations include views on a gay voice for women generally as well as

indications of a development in preferred sexual orientation terminology, particularly among younger people. There will also be a consideration of how the researcher fits within the present study and how researchers should approach linguistic work, particularly when they are not part of the community they are engaging with. The chapter will end with a discussion on how language and sexuality research can move forward in variationist studies but also how language and sexuality research can help develop variationist research generally.

Finally, Chapter 6 will contain some concluding remarks about the thesis as a whole. There will also be a brief discussion on the limitations of the present study and recommendations for future work.

Chapter 2. Literature Review

2.1. Introduction

This review of previous work will attempt to account for several key aspects of this thesis. First will be explanations and definitions of sociolinguistic theories that are employed by this thesis. This will then be followed with an outline of the development of queer linguistics. This will cover both the history of research that may fall under the banner of queer linguistics, as well as the development of queer linguistic theory. Following this will be profiles of the phonetic elements that are important for this study. This will begin with a description of Yorkshire accents, as all the speakers for this thesis are from Yorkshire and have Yorkshire accents. After that, there will be reviews of the specific phonetic features analysed in this thesis: fundamental frequency and fricatives, and in particular the voiceless alveolar fricative, /s/. Finally, this chapter will finish with an overview of the research that specifically focuses on gay female speakers or female sexuality, first, to establish what has been done previously, and then to demonstrate how this thesis works to fill gaps that currently exist in the field.

2.2 Sociolinguistic Theoretic Framework

Before focusing more specifically on language and sexuality research, it is important to consider the sociolinguistic theories this research is based in and that strongly influenced both data collection and analysis.

2.2.1 Approaches to Variationist Sociolinguistics

In describing the different approaches to variationist sociolinguistic research, Eckert (2012) puts forth three "waves". While the waves may imply a chronological development of variationist research, this is not the case. First wave approaches still can be, and are, conducted today in on-going research. These waves simply attempt to explain the way researchers may approach not only data collection, but data analysis when working with variationist research.

The first wave is exemplified by Labov's (1966) *Social Stratification of English in New York City*, with a focus on stratified groups of speakers that may be workings towards or away from the standard dialect, depending on specific social factors. The first wave approaches rely on predetermined social categories that the researcher then attempts to create a representative sample of each category. These social categories are generally based on more macro-level identity factors, such as social class, gender, race, etc. Stylistic variation generally takes the form of formality in first wave research, and there may be a comparison between a person reading from a word list (the most formal) to a sociolinguistic interview that attempts to access the speaker's vernacular form (the least formal).

The second wave shifts the focus to locally salient identities and instead connects vernacular or standard usage to social motivations by an individual. In order to obtain this knowledge of locally salient factors, second wave research employs more of an ethnographic approach to research. One of the key second wave studies is Milroy's (1980) research in Belfast. In her study, Milroy followed different groups of working-class speakers and tracked their social connections to other members of the groups. In her work, and others that follow a second wave model, Milroy asserts that there is more individual agency to the usage of vernacular or standard forms and that certain forms may be used to demonstrate a connection to different communities.

The employment of ethnography was one of the factors leading to the development of third wave approaches. Third wave studies place all the agency with the speakers, much like the second wave, but also allow for more intraspeaker variation than the other waves. In third wave studies, the researchers observe participants and attempt to discover what categories are locally salient to that specific community, instead of relying on predetermined, macro-social categories. Moore (2004) observed that orientation to or away from school was the key factor separating out different groups of school-age girls in her ethnographic study of a high school. She labelled the different groups as Townies, Populars, Geeks, and Eden Village. These groups were all locally salient and did not rely on macro-social categories. While

certain groups may have had more working-class members, for example, it was not their social class that was meaningful in their use of different linguistic features. Another hallmark of third wave research is the belief that a person's identity is not fixed, and their linguistic realisations will change depending on the part of their identity they are putting forward at a given time and context. An example of a study demonstrating the fluidity of identity comes from Podesva (2007) and his research with a participant he identifies as Heath. In this study, Podesva demonstrates that Heath uses more falsetto when he is with his friends during a barbeque and he uses significantly less falsetto when he is speaking with one of his patients or when he is on the phone with his father. These are important insights because is it clear that at all points, Heath is still being himself and all of these versions are "authentic" i.e. he is not acting or performing in some way. However, in these different contexts, certain parts of his identity are more relevant, and therefore he uses specific linguistic features, to bring them to the forefront. In order to understand these different ways to present a part of one's identity, there will be a further discussion of stance, indexicality, and performativity below.

This current project has taken a conscious third wave approach to the research. While some elements had to be shifted to accommodate the change in communities of practice, it was always important to find identity categories that were locally salient to the participants, instead of relying too heavily on predetermined categories. There was a continued attempt to add as much nuance to all the participants as possible, both within the production study and perception study. These third wave approaches will be described in full in the respective chapters of each study. Section 5.3.2 will then have a fuller discussion considering third wave approaches not just in this present thesis, but in variationist sociolinguists as a whole.

2.2.2 Stance, Indexicality, and Performativity

In order to fully develop some of the third wave approach, it is important to understand the theoretic foundations of this approach. The first concept to consider is *stance*. Kiesling (2009, p. 172) defines stance as, "a person's expression of their relationship to their talk (their epistemic stance—e.g., how certain they are about

their assertions), and a person's expression of their relationship to their interlocutors (their interpersonal stance—e.g., friendly or dominating)." He acknowledges that epistemic and interpersonal stances often combine to create a certain type of interaction. For example, if a person is lecturing the person they are speaking to, they are likely demonstrating that they are very knowledgeable about a topic, which is their epistemic stance, but also demonstrating their authority over the interlocutor, which is the interpersonal stance.

In order to undertake variationist sociolinguistic research, typically one must code for these different stances in order to conduct comparative analysis. However, as Kiesling (2009) addresses, it can be difficult to account for every different stance that a person could be taking in the course of an interaction. Instead he describes the use of speech activities to implicitly code for stance. One speech activity that Kiesling accounts for it gossip, which demonstrates a close social relationship between the speakers and a certain level of authority on what they are talking about. Yet even these speech activities can be combined into larger interaction categories consisting of "Social", "Informational", and "Discourse Management".

Holmes-Elliott and Levon (2017) use this view of stance and expand on it in their study of /s/ in two British reality television shows. They not only code their data based on the three interactional categories considered above, but they also account for the idea of *face* based on work by Brown and Levinson (1987). This allowed for the researchers to not only compare data across the categories of social, information, and discourse management interactions, but also divide it further to consider discourse management, non-threatening social, threatening social, non-threatening information, and threatening information. This type of coding adds a level of nuance to the data that follows a third wave variationist approach. Instead of viewing the participants as having fixed identities that are influenced primarily by their gender and social class, this research shows that there is significant intraspeaker variation.

The second concept after stance to consider is *indexicality*. Indexicality refers to the social meanings utterances have that extend beyond just their referential meanings. Johnstone (2016, p. 633) writes, "A sign (a word, a gesture, a glance, a hairstyle, or

10

anything else that can be meaningful) is indexical if it is related to its meaning by virtue of co-occurring with what it is taken to mean." In this way, two different phonological realisations of the same word can invoke different social meanings based on concepts such as class, gender, ethnicity, and much more.

In her paper focusing on language and gender research, Ochs (1991) demonstrates that there are very few instances in English that gender is directly index; this may occur when using gendered pronouns like "he" or "she" or gendered titles like "Mr" or "Ms". Instead, it is more common that gender is indirectly index. Kiesling (2004) shows this process indirect indexicality in his research on the usage of "dude" in the United States. Kiesling argues that "dude" directly indexes concepts of nonchalant comradery and associations with suffer culture. Nonchalant comradery and suffer culture then indirectly index masculinity, which then connects usage of "dude" with male speakers and interlocutors.

Importantly, Ochs (1991) writes on the non-exclusivity of variable features of language and that any speaker, of any identity can use the same feature, but in doing so, may be indexing completely different social meanings. Eckert (2012) cites different researchers who all found that aspiration of intervocalic /t/ was indexing different types of identities based on the groups being studied. Similarly, research from Maegaard and Pharo (2016) shows that fronted /s/ realisations index qualities such as femininity and homosexuality when the speaker is perceived to be from a "modern" group, but fronted /s/ can also index qualities such as "immigrant" and "gangster" if it is from a speaker perceived to be "street". These examples demonstrate the complexity of indexicality, and the fluidity variable features have, depending on the other social markers associated with them.

When considering identity, it is also important to discuss the concept of *performativity*. Performativity is the idea that repeatedly engaging in performative acts, we bring certain identities into being. Performative acts can consist of things such as which toilet we choose to use in public, wearing distinct clothing, wearing make-up, among a myriad of other ways to signal a certain identity. In consideration of gender specifically, Butler (1990, p. 33) argues, "Gender is the repeated stylization of the body, a set of repeated acts within a highly rigid

regulatory framework that congeal over time to produce the appearance of substance, of a natural sort of being". Therefore, there is nothing inherently gendered when we are born, but instead we bring gender into existence by performing certain acts that are associated with a gender.

While this is not an exhaustive list of different theories that are relevant to variationist sociolinguistic studies, stance, indexicality, and performativity are particularly relevant to third wave sociolinguistic research and to this thesis specifically.

2.2.3 Intersectionality

One final concept to consider in this initial discussion on sociolinguistic theories is intersectionality. The term *intersectionality* was originally created by Crenshaw (1989) to describe the unique prejudice faced by Black women while discussing the inadequate legal proceedings to handle this specific type of injustice. Intersectionality theory is the belief that no one identity (and particularly macro-social categories like "female" or "working-class") is enough to fully describe an individual's identity. Instead, one must consider not only the myriad of identities one may have, but, importantly, how these identities interact and intersect to create a unique experience that is dynamic.

Levon (2015, pp. 297–8) describes three core tenets to intersectionality: it is "the product of multiple and intersecting systems of social classification", the intersections are dynamic and created in the moment, and finally "intersections are not to be viewed as 'crossroads' of two or more discrete and already existing categories but rather that intersections are themselves formative of the categories in question". Based on these core beliefs of intersectionality, Levon argues that variationist linguistics should not only be considering macro-social categories to describe, say, how language may be gendered. Instead, researchers need to consider multiple identities that are contextually relevant to the data and how the intersections of these data demonstrate a unique experience for the participant.

The consideration of multiple identity factors, as well as the dynamic view of identity creation and performance, allows intersectionality to fit well within third wave variationist studies. One example of such a study is the work of Podesva and van Hofwegen (2016) who considered not only the macro-social categories of gender, sexuality, and rural/urban identity, but significantly how these interacted in such a way to present specific gendered identities. Views of intersection can also be seen in Moore and Podesva's (2009) work considering the unique combination of gender, social class, and attitudes towards school that contribute to certain tag-question usage. While there are many more studies that effectively use intersectionality in their analysis, these two works demonstrate the effectiveness of this theoretical framework in variationist sociolinguistics.

Intersectionality has been particularly used in language, sexuality, and gender research, as pointed out by Levon (2015) and therefore is especially relevant in this current research. This will be discussed further both in the relevant discussions for each study, as well as the overall discussion in Chapter 5.

Having considered some of the key theories employed in this thesis, there will now be a more focused view on a specific branch of linguist research: queer linguistics.

2.3 Queer Linguistics

While language and sexuality research appears as early as the 1940s, much of this early work would not typically be considered in present language and sexuality studies, as it predominantly focused on gay lexical guides (Kulick, 2000). Generally, the advent of language and sexuality research we recognise today began in the 1990s. Oftentimes this new subfield focusing on sexuality is referred to as *queer linguistics*, though not all researchers necessarily accept this label. The following section will consider the origin of language and sexuality research, as well as presenting an introduction to queer theory and queer linguistics. In order to contextualise the discussion of the word *queer* and queer linguistic research, I will give a brief history of language and sexuality research that has led to the development of the present-day language and sexuality field.

2.3.1 History of Language and Sexuality Research

In his article discussing the scope of language and sexuality research, Kulick (2000) describes in detail the progression of linguistic research on sexuality and the fundamental changes to it over the years. He writes that much of the initial research consisted of word lists and attempts at making dictionaries that defined terms thought to be commonly used by gay speakers, particularly gay men. The year 1981 marked the first full volume devoted to speech and sexuality research with the publication of *Gayspeak: Gay Male and Lesbian Communication* (Chesebro, 1981). While this volume still focused heavily on the vocabulary of gay male and lesbian speakers, there was more consideration of the social and political ramifications of certain speech forms and what using this vocabulary might mean for a speaker. This marks the beginning of viewing LGBTQ speech more fully and was a step towards queer linguistics.

From the 1980s onwards, more research was published that examined the language of both male and female gay speakers beyond their lexical habits and also included studies of speakers that identified as bisexual and transgender. The 1990s saw the publication of *Beyond the Lavender Lexicon: Authenticity, Imagination, and Appropriation in Lesbian and Gay Languages* edited by Leap (1995), and *Queerly Phrased: Language, Gender, and Sexuality* edited by Livia and Hall (1997). Both of these publications moved away from vocabulary and considered a greater range of communication modes, including sign language, pitch perception, and coming-out stories. This introduced research in the field that focused on how a speaker can use language to index their sexual orientation, including how this might sound and the social ramifications of these speech qualities.

One study that considered how speakers may index a LGBTQ identity and that exemplifies queer linguistics is Queen (1998). In her paper, Queen (1998, p. 203) outlines some of the conversation strategies repeatedly seen in research on male and female gay speakers as follows: "the rhetorical use of assumed 'shared' cultural understanding, parodic appropriation of stereotyped, gendered language use, *covert communication*, and *co-narration* among others" (emphasis in original text). While she acknowledges that these habits may be evident in other communities of

speakers, it is how they are employed by specifically LGBTQ groups that designates them as LGBTQ markers. In her 1998 study, Queen recorded a group of speakers that mostly knew each other, but had fairly loose connections and did not form a community that interacted regularly. She argues that because these participants cannot rely on common knowledge that they have by virtue of being familiar with one another, they instead partially employ a sense of shared community by all identifying as LGBTQ and using the strategies listed above in order to build this sense of a shared LGBTQ identity. Queen's study demonstrates that there may be aspects of language used to index a person's sexual orientation that span further than a specific, local community.

As Queen (1998) noted above in her list of strategies employed by LGBTQ speakers, one aspect of LGBTQ speech is the use of stereotyped, gendered language, typically used with a sense of humour or parody. Gendered speech has regularly been associated with language and sexuality research; stereotypical speech assumed to be used by a gay speaker is thought of as somehow differing from speech typically associated with that person's gender. As Cameron and Kulick (2003, p. 6) write, "…homosexuality is very commonly understood as *gender* deviance… Gay men are commonly thought to be effeminate (hence such insulting epithets as English *pansy*), while lesbians are assumed to be 'mannish' or 'butch'" (emphasis in original text). This concept of gay speech diverging from speech typical of one's gender is the crux of sociophonetic research on sexual orientation. Popular opinions among non-linguists correlate a stereotypical gay male voice with phonetic productions that are more closely aligned with those of straight women than with those of straight men. While there are fewer stereotypes of female gay speech, it is commonly thought to be more masculine, and is therefore associated with straight men.

This idea of gay speech being a "deviation" from speech normally associated with the binary view of gender has led researchers to focus on phonetic features that are commonly correlated with gendered productions. As Pierrehumbert *et al.* (2004, p. 1906) write, "A baseline for understanding differences between GLB [gay, lesbian, bisexual] and heterosexual people is provided by general speech-production differences between adult males and females. These occur as a function of both

anatomical differences and social factors". By researching how male and female gay speakers produce features that are stereotypically gendered, researchers can understand whether gay speakers are using these features in a unique way to index their sexual orientation.

2.3.2 Use of Queer

> "The use of *queer* here is intended to convey any variety of lesbian, bisexual, transgender, or gay culture/society and is otherwise purposefully left vague in terms of who actually claims membership in or is considered to belong to such a culture/society" (Barrett, 1997)

Before it is possible to examine queer theory or queer linguistics, it is vital to first address the use of the word *queer*. There are two different perspectives when considering the use of the word *queer* that will be important for the present study. First, there will be a consideration for how *queer* is used outside of academia. This will consider the origins of the term, the pejorative usage, and then a discussion on its reclamation. Following this, there will be a consideration of how *queer* is used in an academic sense and the connection this has to queer theory.

Queer has been used as a derogatory slur for homosexual people, and in particular gay men, for over a century; however, this is not how the word originated. Brontsema (Brontsema, 2004) cites the origin of *queer* as "general non-normativity separable from sexuality", with Oxford English Dictionary citations from as early as 1621. It was not until the beginning of the twentieth century that *queer* was used as a pejorative term for homosexual men. Partridge (1967) records the first instance of *queer* to mean homosexual as 1937 in the USA, while the Oxford English Dictionary (2018) has examples from as early as 1914.

However, in the late 1980s there began a movement towards *queer* being used in a more neutral or even positive way, particularly by people who identified as LGBTQ. This marks the linguistic reclamation of *queer*. Chen (1998, cited in Brontsema 2004, p.1) says reclaiming "refers to an array of theoretical and conventional interpretations of both linguistic and non-linguistic collective acts in which a

derogatory sign or signifier is consciously employed by the 'original' target of the derogation, often in a positive or oppositional sense." Brontsema (Brontsema, 2004) describes the first public reclamation of the word *queer* through the creation of "Queer Nation, an off- spring of the AIDS activist group AIDS Coalition to Unleash Power (ACT-UP)". She outlines that *queer* was seen a more inclusive than *gay/lesbian*, as it covered a wider range of identities.

Now, 30 years on, *queer* has become much more widely used by the LGBTQ community in a positive manner to include all people that do not identify as cis-gendered or heterosexual (Zosky and Alberts, 2016). However, as there continue to be instances of *queer* being used in a derogatory manner, there are still those within the LGBTQ community that do not feel comfortable with the use of *queer*, even as a reclaimed word. Through my own experience working with LGBTQ charities, some members were not comfortable with the use of *queer* because they had negative memories attached to it. For this reason, some groups may still avoid using *queer* if possible, so as to not risk offence to anyone in the community. In their study of how participants respond the word *queer*, Zosky and Alberts (2016) found that there is a complicated relationship and the majority of their participants accepted *queer* as self-referral, but were hesitant to refer to someone else using this term.

Along with the discussion above, there is also an academic, theoretical definition of *queer* that is significant when considering queer theory and queer linguistics. Barrett writes that the academic use of *queer* was meant to "reclaim (and hence redefine) a pejorative term so that it has no referent, but is a purely indexical sign in both form and use" (2002, p. 27). In this, Barrett asserts that by taking back the word *queer*, academics have attempted to recreate it in a way that has no certain meaning that can be mapped on to any specific group of people. By having no defined or clear referent allows *queer* to be completely inclusive of all people in research. However, for some, this definition of *queer* leaves it too vague and therefore open to confusion or debate and others argue that while academics may assert that *queer* does not signal any certain group, this is not what happens in practice.

Arguably, this sense of *queer* meaning "non-heterosexual" has leaked into the academic understanding as well. Livia (2002) acknowledges in her chapter of *Language and Sexuality* that the first issue with the word *queer* can be its inherent vagueness. While many do not have a clear definition of how they classify *queer*, arguably it could be used as an umbrella term that is "simply a useful shorthand for people who are not straight" (Livia, 2002, p. 87). Echoing this concern, Kulick (2002, p. 66) argues "if queer is not the same as lesbian, gay, bisexual, or transgender – as all queer theorists insists that it is not – why, then, is the only language ever investigated to say anything about queer language the language of people who self-identify, or who researchers believe to be, lesbian, gay, bisexual or transgendered?".

If it is unclear who falls under the category of *queer* and it is arguably used in a way to encompass anyone who is not heterosexual, there is also a risk of creating a category that is overly large. Livia (2002, p. 87) writes that some communities that may be seen as *queer* are "transsexuals, sex workers, gay men, and lesbians" and then we must ask the question, are these groups inherently connected in such a way as to fall under the same branch of research? Are there shared values and commonalities amongst these groups such that they should be categorized together under one heading as *queer*?

With this broad usage of *queer*, Leap (2002, p. 61) argues that it will not aid researchers in addressing the very real issues of "discrimination, homophobia, racism, sexism, class privilege, and other themes which structure conditions and languages of everyday experiences for so many women and men (same-sex oriented and otherwise) in late modern society". He argues that by not acknowledging the differences amongst those in the queer community and simply focusing on "sexual/gendered transgression as the cornerstone of queer experience" the dominant subject of queer linguistic research in a Western context will continue to be white, middle class, gay men (2002, p. 49).

Though several authors in *Language and Sexuality* (Campbell-Kibler *et al.*, 2002) highlight issues with the use of the word *queer* that are important to consider, this publication is now 17 years old at the time of writing and *queer* has continued to

develop in its usage and acceptance. In a more modern usage, *queer* is used in a way that is inseparable from the connection to queer theory, and instead refers to challenging hegemonic norms. For example, in his introduction to the volume *Language and Masculinities*, Milani (2015) uses the word *queer* to describe how the book challenges traditional masculinity by including research that includes gay women and transgendered and intersex people. In this way, the book is queering masculinity.

This newer usage of queer to describe challenging norms is arguably only an academic usage. In his article reviewing queer linguistics, Motschenbacher writes, "…the anti-normative interpretation of *queer* is mainly restricted to the academic debate, whereas outside academic contexts *queer* is usually used in the sense of 'non-heterosexual' and has therefore gone through a process of identity materialization" (2011, p. 152). Here Motschenbacher presents the idea that academics continue to use *queer* in a way that does not have a referent the way Barrett presented above, as well as the norm-challenging way, but also acknowledges that outside of academia, in the general population, *queer* is used to describe anyone that does not fall under the labels straight and cis-gender.

This use of *queer* as both an umbrella term for LGBTQ people and as a challenge to hegemonic norms can make deciphering the meaning difficult at times. Many recent publications use *queer* in the title (Hall, 2013; Milani, 2013; Motschenbacher and Stegu, 2013; Zimman, Davis and Raclaw, 2014) and at the 2018 Lavender Language and Linguistics Conference, it was used in the titles of over ten papers as well as in the names of two panels. It may be unclear how an individual researcher interprets and uses *queer* in their work, and if they are challenging norms or simply referring to LGBTQ people. However, an understanding of queer linguistic research can aid in understanding of *queer* and how it may be used.

2.3.3 Queer Theory and Queer Linguistics

Queer linguistics originated from queer theory, which was a general approach to research in the humanities and social sciences. As Motschenbacher & Stegu (2013) note, queer theory began in the USA in the 1970s and 1980s in response to the gay

and lesbian rights movements. While there is no one single definition of queer theory and many have interpreted queer theory in different ways, two of the main aspects of queer theory are that it questions hegemonic norms and uses gender and sexuality as starting points for research and analysis (Kirsch, 2000). In connection to questioning various societal norms, queer theory rejects the dichotomy between heterosexual and homosexual and demands a more complicated approach to sexuality. Queer theory can be connected to poststructuralist thinking and is often associated with work by Butler (1990) and Foucault (1978). Generally, the goal of queer theory is to look critically at heteronormative ideals.

Barrett (2014, p. 196) writes that "[one] of the main goals within queer theory has been to challenge essentialism in various forms, emphasizing instead the ways in which identities and normative ideologies are socially constructed". Queer theory demonstrates that essentialism will hurt those individuals who do not subscribe to the heteronormative standards and therefore queer theory seeks to expand from this essentialism in order to consider all of humanity. One way to move away from these norms is to shift the focus of research onto frequently marginalised groups of people that are not commonly featured in academic study.

This is the reason for using gender and sexuality as a starting point for research. Barrett (2002, p. 25) summarises the goals of queer theory exceptionally well when he writes, "the goal of such research is not intended simply to increase understanding of 'queer' behaviour, but to increase understanding of human behaviour and to question exclusionary theoretical assumptions across academic disciplines". Therefore, queer theory does not seek only to include a more diverse range of research, but to have academia as a whole question the theoretical frameworks that have been used previously that are overly reductive and therefore do not consider all of humanity.

In an attempt to distance itself from hegemonic ideals, queer theory seeks to examine the ways in which individuals perform aspects of identity, in particular gender and sexuality, and how these performances may be regulated by society. However, queer theorists acknowledge the instability of these performances and the fluidity of how an individual can present certain aspects of their identity. Berlant

and Warner (1995, p. 344) write, "Queer publics make available different understandings of membership at different times". Identity is created in the moment.

One aspect of these identity-performances researchers may consider is language, and how speakers may use it to construct and perform a certain identity connected to gender and/or sexuality. This connection between performance, queer theory, and language has led to the development of queer linguistics.

Motschenbacher (2011, p. 153) explains how queer linguistics differs from previous views of variationist sociolinguistics:

> The relationship between language and identity is here rather considered as constructive, i.e. identities are seen as constructed in the very moment of language use. In other words, what used to be taken as a fixed grid of identity facets to which speaker behaviour may be correlated is in a post-structuralist theorization conceptualized as an intrinsically unstable, procedural accomplishment that may orient to such dominant identity notions but at the same time negotiates them contextually in a process of fluctuating identity constructions. Identity categories like 'woman' or 'man', for example, would no longer be treated as self-evident, stable demographic variables in such a theorization.

Based on the previous discussion of the waves of variationist sociolinguistics (Eckert, 2012), it is clear that the distinction Motschenbacher is making between queer linguistics and variationist linguistics is more based on first and second wave variationist research. The view of identities being unstable and created in each moment matches the tenets put forth in third wave variationist studies.

An example of queer linguistics in practice can be seen in a paper by Sauntson and Morrish (2012). In their study of a women's university football team, the authors write that they have followed queer theory practices in their analysis by not assuming that identity is fixed and instead viewing it as flexible. The researchers demonstrate that the speakers did index their sexual orientation in the conversation,

but that these tools to demonstrate their sexuality were all contextual and specific to an understanding shared by this particular community. These same markers may not be valid or even understandable to other speakers outside the community and outside of this particular setting.

This view of identity factors as unstable and ever-changing depending on context is the cornerstone of queer linguistics, but is also what makes the research particularly challenging and the existence of queer linguistic research at times contentious, particularly for those who do not work under the queer theoretical framework. Queen (2002, p. 70) wrote in the early 2000s, "…most of the work that gets placed under the label 'queer linguistics' is not specifically queer theoretical but rather based on data from queer subjects". This is to say that previous work that may be seen as queer linguistics because of who was researched may not have taken queer linguistic approaches if the research assumed a level of fixed identity among the participants or perpetuated hegemonic norms. This distinction is important when considering the difference between queer linguistics and, more generally, language and sexuality research.

Not everyone that works on language, gender, and sexuality research uses a queer linguistic approach. Some of the early work, particularly dictionaries of "gay language" as described in the previous section, would not be categorised as queer linguistic research, as there was no challenge to hegemonic norms and these dictionaries still employed an essentialist view. While it would not be expected that the work in 1940s would fall under queer linguistics, this example is given to illustrate the difference between language and sexuality research and queer linguistic research. As acknowledged with the word *queer*, it is important to distinguish between the umbrella usage of *queer*, which may only specify language sexuality research, and *queer* in connection to queer linguistics.

Queer theory and queer linguistics continue to develop and there are still many researchers considering the merits of a queer analysis. In 2019, a special issue of *Language in Society* specifically considered queer linguistics and how to approach anti-normativity. Researchers such as Wiegman & Wilson (2015) have questioned queer theory's anti-essentialist tenets with how many queer theorists address anti-

normativity. This spurred an in-depth publication on how sociolinguistics can uniquely address continued queer theoretical developments (Hall, Levon and Milani, 2019). While this special issue is not the only research continuing to engage with queer theory, it stands as an example of the ongoing work done within the field and the shifting ideologies within the theory.

Though the present thesis does not employ a strict working of queer linguistics or queer theory, it is important to understand the basic concepts when considering the entire project. Queer linguistics has been influential in how the data has been analysed and is frequently considered in the work that inspired the current research. A further discussion of the theoretical implications of the thesis will be presented in section 5.3.2, after the data has been presented.

2.4 Profile of Yorkshire Dialect

Having considered some of the theoretical framework that influenced this research, it will now be helpful to consider the phonetic landscape of the studies. The present research consists of two separate studies, both of which focus on Yorkshire speakers. For this reason, it is important to first consider the typical features associated with a Yorkshire dialect more broadly.

Yorkshire is part of the north of England and therefore typically follows common Northern English patterns. A map of Yorkshire can be seen in Figure 1. Two of the most recognisable features of a Northern accent are a FOOT-STRUT merger and BATH vowel realisations (Wells, 1986). A lack of FOOT-STRUT distinction indicates that these vowels are not distinguished and therefore *put* and *putt* are realised as homophones and can broadly be transcribed with /ʊ/. One of the other significant features of northern English is the BATH vowel is generally realised as /a/, opposed to the South, where the BATH vowel is realised as /ɑ:/.

Figure 1 Map of the British Isles with Yorkshire highlighted

Yorkshire accents also have monophthongal realisations of both FACE and GOAT vowels, transcribed as /e:/ and /o:/, respectively. While research from Haddican *et al.* (2013, p. 396) show that there are signs that these two vowels are undergoing changes towards diphthongisation, they write that "participants frequently identified monophthongal realizations of FACE/GOAT as typifying York or Yorkshire dialects". The researchers observed that realisations of FACE and GOAT were strongly correlated with how participants orientated to the community; participants that had a more positive connection to the area used more monophthongal realisations.

Wells (1986) also writes of the "Yorkshire Assimilation" which occurs when a final voiced obstruent is realised as voiceless when it comes into contact with an initial

voiceless obstruent. The example given by Wells is the realisation [ˈbɛttaɪm] for *bed-time*.

Finally, northern dialects generally, and Yorkshire dialects in particular, are associated with definite article reduction (DAR) (Jones, 2002). DAR refers to a vowel-less realisation of definite articles, and in some cases zero realisations of definite articles. This leads to typical orthographic transcripts such as "down t'pub". This can also commonly be seen on merchandise sold in Yorkshire. An example of this can be seen in Figure 2, which depicts a mug sold on a website titled *Utterly Yorkshire*.
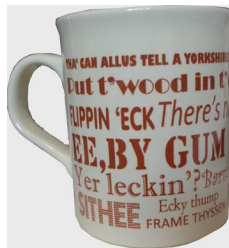
Figure 2 Novelty mug depicting Yorkshire dialect

There is variation within Yorkshire and locals are typically aware of the variability between different regions in Yorkshire. Cooper (2019) shares examples of comments from his survey of Yorkshire dialects that highlights participant awareness of differences within the region and write in the need to specify sub-regions. Cave (2001), Cooper (2019) Finnegan (2015), Llamas (2015), and Watt (2013) are just some examples of research that demonstrates variation within Yorkshire. However, much of this variation is based on vowels or lexical differences and the present study does not focus on vowel realisations. For this reason all the participants considered together under the general region of Yorkshire.

Along with considering some of the features of a Yorkshire accent, it is worth briefly considering some of the social identities and perceptions associated with Yorkshire and the dialect. Wales (2000) describes the perception of the north and south of England as binary oppositions. She includes some of the oppositions as: "hot" and "cold", "old" and "young", "rich" and "poor". Not only is Yorkshire associated with the common northern stereotypes, but the county has its own set of

stereotypes as well. Fletcher (2012, p. 230) describes Yorkshire as the following: "Yorkshire's industrial cities, hard work ethic, tendency towards self-deprecation, rustic charm, overt masculinity, suspicion of strangers and, above all else, pride, capture much of what the region is about."

The Yorkshire dialect is one that many in the UK are aware of and hold certain associations with. Cooper (2019, p. 68) describes the Yorkshire dialect as being enregistered "and indexes social values including the geographical location of Yorkshire along with more abstract concepts like 'plain speaking', 'authenticity', 'independence', 'sense of humour', 'geniality' and 'hospitable". In his survey considering the features commonly associated with the Yorkshire dialect, Cooper (2015) found that not only were participants aware of a Yorkshire dialect, but they highlighted several features that they associated with Yorkshire, including: DAR, use of *nowt* ("nothing"), use of *owt* ("anything"), use of *summat* ("something"), h-dropping, and more.

These features of a Yorkshire dialect and some of the societal views of a Yorkshire dialect have been highlighted to explain the benefits of working with this dialect specifically for the present study. This is a dialect that surveys have demonstrated people are not only aware of but have enough of an awareness that they have opinions on different aspects of the dialect. This makes it a strong dialect to frame the perception study (described in Chapter 4) around, as British people are already primed to consider the Yorkshire dialect and they are familiar with it. This would be a plausible study that they would engage with, and hopefully by focusing more on the prompted dialect, they would not consider the questions on sexual orientation and femininity as closely – the benefits of this are describe further in Chapter 4.

While this section cannot provide an exhaustive description of the Yorkshire dialect or the perception of Yorkshire people, this section acts as an introduction to some of the important qualities that are present later in the thesis. All the participants in the production study come from various parts of Yorkshire and use the phonetic features described above to differing degrees. The speaker that provided the stimuli for the perception study is also from Yorkshire and some of these features were likely identifiable to many of the participants in that study. Having described a broad view

of the dialect of many of the speakers in this thesis, it is important to consider the specific phonetic features that are analysed in both studies.

2.5 Phonetic Variables

The following section describes the two phonetic variables that are considered in this thesis: fundamental frequency and /s/. First, the variable will be described from an articulation view in order to understand how these sounds are produced. Then there will be consideration of how the variable interacts with gender, based on previous research. Finally, there will be a specific focus of the variable and previous sexuality research.

2.5.1 Fundamental Frequency

While the terms are sometimes used interchangeably, pitch and fundamental frequency (henceforth F0) refer to two different aspects of the speech signal. F0 is the measurement of voicing during an utterance and pitch is its perceptual correlate. Pitch refers to an auditory evaluation of a sound, while F0 describes the production of a sound. F0 is a measurement of the number of opening and closing cycles of the vocal folds in a second, notated in Hertz (Hz) (Ladefoged and Johnson, 2011). Due to the biology of the vocal tract, there are some physiological constraints on F0. Speakers have an F0 range that they utilise in their normal speaking voice which is based on what they are physically comfortable producing; this range is their modal voice, which involves regular vibrations of the vocal folds. Variation in the function of the vocal folds (i.e. tension, incomplete closures) will affect F0 production. For example, a speaker may place their vocal folds under increased tension so only a small portion of each is able to vibrate, causing creaky voice or vocal fry, or they may keep their vocal folds slightly apart during voiced sounds creating breathy voice or murmur. This thesis will be primarily focused on modal voice and a person's typical F0 range.

Sound energy begins at the vocal folds when air passes between them and the folds vibrate; the air is then expelled largely through the lips. This energy is influenced by the length and shape of the vocal tract and it is this energy that listeners hear which they then perceive as pitch. There are different biological factors that

influence F0 and some differences can be based on sex. According to Simpson (2009), male vocal folds tend to vibrate more slowly due to being longer and thicker than female vocal folds. This causes men to generally have a lower F0 than women. Simpson writes that the average F0 for male English and German speakers is about 100-120 Hz and for women the average is roughly 200-220 Hz. Traunmüller & Eriksson (1995) found the average F0 for men to be 120 Hz for men and 210 Hz for women. Both studies cite previous research when addressing the average F0 based on sex.

While F0 is correlated with sex, it can be altered and influenced by other factors, such as gender identity. Work with transgender speakers, particularly transgender women, has shown that F0 may be deliberately altered by the speaker in order to convey a particular gender identity (Brown *et al.*, 2000; Gelfer and Schofield, 2000). In these studies, as well as that of Munson (2007b), higher pitch was associated with femininity and lower pitch with masculinity, based on perception rankings by listeners. With the correlation of F0 with sex, as well as the fact that it may be influenced by gender identity, F0 and pitch are commonly studied in relation to sexual orientation.

Another aspect of F0 that is commonly considered in studies of gender and sexuality is F0 range. The range describes the lowest modal frequency used by the speaker to the highest point. If a speaker is described as *monotone*, then they are using a small F0 range and staying relatively "flat" in their F0 range. Some believe that women typically have a larger pitch range than male speakers (Lakoff, 1973; Berryman-Fink and Wilcox, 1983), which has led to this feature being commonly studied in language and gender as well as language and sexuality. However, as Lennes *et al.* (2016) write, this is not necessarily true. When pitch range is measured on an absolute Hertz scale, women do appear to have a larger range than men. However when range is measured in semitones, that "difference more or less disappears" (Lennes *et al.*, 2016, p. 36).

Semitones are based on a logarithmic scale corresponding to musical intervals. A difference of 12 semitones is one octave and each octave corresponds to a doubling of frequency (for example 200 Hz is one octave higher than 100 Hz and 800 Hz is

one octave higher than 400 Hz) (Thomas, 2011, pp. 59–60). The semitone scale has been argued to be a more accurate representation of how listeners hear changes in F0 (Waksler, 2001).

Esling and Edmondson (2010) clarify that when pitch range is considered in Hertz, it appears that women frequently have larger ranges. Yet when pitch range is considered in semitones there is often no difference between the genders. However, it is important to recognize that due to the persistent belief that female speakers have a larger pitch range than male speakers, because of the F0 measurements, pitch range is also frequently considered in sexuality studies. Therefore, in order to connect to the previous research, this thesis also considers F0 range and will provide evidence both in absolute Hertz and semitones.

2.5.1.1 F0 and Sexual orientation

The mid-1980s and early 90s saw the emergence of research considering F0 and LGBTQ speakers. Much of the early work focused on the speech of men and was based on perception of F0 as well as production.

Gaudio (1994) recorded 8 speakers (4 straight men and 4 gay men) reading text passages that were then played in parts to listeners. Participants were asked to judge the speaker's sexual orientation based on his voice. Guadio found that there was no correlation between higher F0 and the ratings of sexual orientation by the listeners. Munson (2007b) also did not find any significant correlation between sexual orientation and average F0 when studying men and women. Instead, his study found that average F0 was more closely correlated with perceptions of masculinity and femininity than with gender. Smyth and Rogers (2003) write that of the numerous phonetic features they have studied in association with sexual orientation in men, mean F0 does not have a significant correlation with perceptions of male sexual orientation by listeners. This previous research has shown that listeners do not consistently perceive male speakers with a higher F0 to be gay, and therefore pitch is not the only factor that listeners use when judging a male speaker's sexual orientation.

Moonwomon-Baird (1997) investigated the influence of female sexual orientation on pitch by researching conversations between two pairs of speakers – one conversation between two gay speakers and another between two straight women. She concluded that the straight women had more variation in the pitch they used in the 2.5-minute conversation and also that their pitch was generally higher than that of the gay speakers. However, as Moonwomon-Baird had only four participants, others have attempted to recreate elements of this investigation to test its accuracy, as we shall see below.

Instead of relying solely on the perception of listeners, other studies have compared F0 productions between gay and straight speakers in order to determine whether sexual orientation has an effect on F0. Waksler (2001) found no significant differences in pitch range between gay and straight female participants based on her study of 12 straight women and 12 gay women retelling the story of the *Wizard of Oz*. However, Van Borsel, Vandaele, and Corthals (2013) found in their study of 34 gay women and 64 straight women that the gay women had a significantly lower average F0 than the straight women, as well as significantly lower pitch variation. Rendall, Vasey, and McKenzie (2008) found that F0 was inconsistent based on sexual orientation; the female speaker with the highest average F0 identified as gay, but the five speakers with the lowest average F0 all also identified as gay. Due to the extensive variation within the groups of gay and straight female speakers, there was no statistically significant difference in F0 based on sexual orientation.

Previous research has shown that there are no conclusive results to show that a connection can be drawn between sexual orientation and average pitch or pitch range, either in production or perception. Findings thus far have been inconsistent for both male and female speakers. For this reason, pitch should continue to be studied as a potential correlate of sexual orientation, but other phonetic markers must also be considered and investigated. Findings related to F0 from my current study will be presented later in this thesis.

2.5.2 Fricatives

Schwartz (1968) determined that the sex/gender of an English speaker could be accurately distinguished based on isolated voiceless alveolar fricatives (i.e. tokens of /s/) and that women tend to have higher peak frequency in /s/ than men. Stuart-Smith *et al.* (2003) discovered that of the variables they considered, only /s/ showed a clear and consistent gender distinction, with women having an overall higher mean and peak; both of these measurements are based on the midpoint frequency of the spectral curve. It is important to note that higher measurements did not apply to a specific group of participants in the Stuart-Smith *et al* study and these participants will be discussed further below. Podesva and Van Hofwegen (2014) write that as men tend to have longer vocal tracts than women, /s/ realisations tend to be lower in frequency for men than women, causing men to have a lower spectral centre of gravity than women. Therefore, data has consistently shown that /s/ can be used to signify a speaker's gender.

Productions of /s/ regularly show a significant gender distinction, and this makes it an ideal segment to study with respect to language and sexual orientation. If a token is generally realised in a certain way for a particular gender, it may be telling if a person of that same gender but of differing sexual orientation varies in their production from the statistical norm for their gender. And if they were to vary, how great, and of what sort, is the variation? As most of the previous research has been conducted without considering sexual orientation, it is important to fill in these gaps and understand the patterns in /s/ when sexual orientation is the variable of interest.

Another reason /s/ is an ideal variable when considering language and sexual orientation is that research has shown that /s/ may be modified in order to index certain identity factors for a speaker. The Stuart-Smith *et al* (2003) study cited above investigated /s/ realisations in connection with sex, gender, and social class. They found that the /s/ realisations of working-class girls in Glasgow aligned more closely to those of male speakers than those of other groups of female speakers. Working-class girls had a higher spectral skew with values closer to 0 and closer to the male range, as opposed to working-class women, who had negative values. If

this group of female speakers is signalling identity factors, particularly gender and class, by producing characteristic /s/ realisations, it is plausible that other groups of speakers may also show signs of variation in /s/ to signal other identity factors.

2.5.2.1 Fricative and Sexual orientation

The study of /s/ has also been a cornerstone of sexuality research in linguistics due in large part to the stereotype of the "gay lisp". Munson and Babel (2007) discuss the prevalence of the gay lisp in popular culture, from television shows to books and stand-up comedy. This stereotype has in itself led to a considerable amount of linguistic research investigating /s/, in both production and perception, in connection with sexual orientation. The stereotype of the gay lisp is associated specifically with gay men and it is partly for this reason that much of the previous linguistic work on /s/ and sexual orientation has focused on male speech. Crist (1997) studied the duration of onset consonants over two experiments in the speech of six participants – two self-identified straight men and four self-identified gay men. He found that the friction in /sp/ and /sk/ clusters had a longer duration in the gay male voices than in the straight male voices. Smyth and Rogers (2003) also found that /s/ and /z/ had a longer duration and a higher peak frequency for gay-sounding men. More recently, however, there has been a wider range of phonetic research that also considers gay female speakers with respect to /s/ production.

Strand and Johnson (1996) synthesised a /s/ to /ʃ/ continuum and found that participants' perceptions of /s/ or /ʃ/ were significantly influenced by visual cues that presented either a male or female face while individual stimuli were heard. Though the same stimuli were played to all the participants, those sound clips that were paired with a male face were more likely to be perceived as /s/ and those paired with a female face were more likely to be perceived as /ʃ/. Munson, Jefferson, and McDonald (2006) studied perceptual differences in the continuum of /s/ to /ʃ/ and also found that productions which were acoustically nearer a canonical /ʃ/ were heard as /s/ when the voice was perceived to be that of a man, but as /ʃ/ when believed to be that of a woman. They also observed that /s/ produced by voices that were perceived to be those of gay women were more likely to be perceived as /s/ than voices perceived to be from straight women. Women believed to be gay were

heard, therefore, as being more masculine. This work shows that sexual orientation may influence at least the perception of /s/.

In a study based in Northern California, Podesva and Van Hofwegen (2014) studied the centre of gravity of /s/ in relation to several factors: gender, sexual orientation, and whether the participant was more country- or town-oriented. They found that straight country men had the lowest average centre of gravity and straight town men had the second lowest average. Straight town women had the highest average centre of gravity, with straight country women having the second highest. In Podesva and Van Hofwegen's study, straight speakers set the poles of the continuum for centre of gravity in productions of /s/ and LGBTQ speakers made up the middle. This demonstrates that it is possible that the LGBTQ speakers in the community were indexing their sexual orientation or gender identity through their pronunciation of /s/.

A final study of /s/ production in connection to sexual orientation to consider is work done by Saigusa (2016). This study examines the effect the addressee has on the speaker and the centre of gravity of the speaker's /s/ production. In the study, two interviews with Jane Lynch, an openly gay actress, were analysed. The first interview was by an openly gay news reporter and the second was with two female TV hosts considered not gay by Saigusa. (The other hosts may not identify as straight, but as they are married to men it is unlikely they identify as gay.) Saigusa found that Lynch had a lower centre of gravity in /s/ when speaking with the openly gay news reporter than when speaking to the non-gay TV hosts. It was also found that when Lynch was discussing LGBTQ issues she had a lower centre of gravity than when she was discussing other topics, such as beginning her acting career.

Though there is no equivalent of a gay lisp for gay women in popular culture, it would be logical to also study the features of the production of /s/ with gay women if it is being used to index social categories such as gender and social class.

2.6 Gay Female Speakers in Linguistic Research

Having considered some of the main sociophonetic features researched in language and sexuality research, this section will focus more specifically on gay female speakers. While gay male speakers have dominated research on language and sexuality, there has also been a significant amount of research on gay female speakers. However, it is important to note that when looking closely at a narrow demographic of speakers, one runs the risk of overgeneralizing or implying that all speakers in that group use certain features or strategies. The research summarized below, as well as the thesis as a whole, do not attempt to say how all gay women globally speak, but instead confines their conclusions to a narrow scope of speakers that have taken part in sociolinguistic research.

As Queen (1998) and Jones (2018) both note, there are certain strategies that have been seen repeatedly in the research on lesbians in sociolinguistics. These similarities amongst speakers have been seen primarily in discourse analysis and will be discussed further below, as discourse and conversational analysis have been one of the primary fields of linguistics to research lesbian speakers. The focus then turns to the sociophonetic research that has included lesbian speakers, as this thesis adopts a sociophonetic approach. Finally, I will demonstrate where the research falls short when attempting to understand gay speakers and aim to show how this current thesis addresses some of these gaps.

2.6.1 Discourse Analysis

There has been a great deal of research that uses discourse analysis in order to understand how female speakers may index their identity within specific conversations and certain contexts. Jones (2018) outlines some of the key initial work on gay speech, which started in the 1980s. All of these initial studies use discourse or conversational analysis to look closely at how self-identified gay speakers may use language in such a way as to index their sexual orientation. Jones demonstrates that many of the studies of how gay women may index their sexual orientation demonstrate stances the speakers take to show their membership of a gay

community. These stances may include the speakers mutually agreeing that homophobia is a problem and sharing their knowledge of LGBTQ history, as in Queen (1998), or the use of humour and teasing around hegemonic female norms, as in Bland (1996, as cited in Jones 2018)

Discourse analysis continues to be one of the common fields of linguistics that researches how female sexuality is expressed. Sauntson and Morrish (2012) undertook a close analysis of a conversation between a university football team whose members identify as both gay and straight. In this conversation the young women frequently talk about sexual orientation and their understanding of how to identify or display sexual orientation. The authors write that this conversation demonstrates how fluidly the speakers view sexual orientation and how they use things like humour to undermine staunch norms.

Shrikant (2014) also uses discourse analysis, and in particular membership categorisation analysis, to demonstrate how a group of speakers index their sexual orientation. Seals (2016) also uses discourse analysis to look at the humour used by a gay female comedian and how she interacts with her audience. While these recent studies are not an exhaustive list of the on-going discourse analysis research in connection to female sexuality, they demonstrate the frequency of discourse analysis within this field of research.

2.6.2 Sociophonetic Research

There has been less sociophonetic research on gay female speakers than there has been with gay male speakers. This may be due, in part, to the fact that there are far fewer stereotypes of lesbian speech than there are of gay male speech. As discussed above there are many references in popular culture to a "gay lisp" or a "gay voice" and it is even prominent enough that entire documentary titled *Do I Sound Gay* (Thorpe, 2014) discussing the topic of a gay male voice was released. However, there is not a similar phenomenon with gay female speakers and there are rarely distinct features that are identified as being particularly "lesbian". Therefore, there are not as many studies that consider specific features of gay female speech. The

following section will revisit the few studies that do consider lesbian speech and consider the overall patterns they may show.

As the studies of F0 and pitch were discussed above in Section 2.5.1, I will not go into great detail about the following research. Instead I will highlight the key findings to then examine patterns later. Moonwomon-Baird (1997) published initial findings of female gay speech and F0 range. Waksler (2001) later developed this F0 range research with her study in San Francisco, California.

Following Waksler's study was an even larger-scale study conducted by Van Borsel, Vandaele, and Corthals (2013). In their study of Dutch speakers, 102 participants were recorded – 34 lesbian speakers and 68 straight speakers. They found that the lesbian speakers had a significantly lower mean fundamental frequency than the straight speakers and that the lesbian speakers had a significantly lower pitch range than the straight speakers. The researchers do acknowledge that the lesbian participants were much more likely to be smokers than the straight participants and wonder if this potentially influenced the results.

One final fundamental frequency to consider is that of Queen (2006) who looked at the characters on the popular US television show *Ellen*. In her research, Queen found that the lesbian characters had a lower average pitch than the heterosexual characters. While it is difficult to understand how natural this may be or how much was influenced by portraying a character broadcast to a large audience, it may be significant in how people expect lesbians to sound – or at least people who watched *Ellen*.

As highlighted above, the other vocal features commonly studied not only in language and sexuality research, but with female gay speakers specifically, are fricatives. As all the studies that considered gay speakers were discussed at length in Section 2.5.2 above, I will not go into exhaustive detail again here. However, it will be helpful to highlight the key findings from these studies.

Munson, Jefferson, and McDonald (2006) considered the continuum of /s/ to /ʃ/ and found that women who were identified as being lesbian or bisexual by participants

were more likely to be perceived as using /s/ instead of /ʃ/ and the productions of fricatives by lesbian and bisexual women had lower centres of gravity than those produced by straight women. Podesva and Van Hofwegen (2016) found similar centre of gravity results in their study of urban and rural speakers in Northern California. In their study, the lesbian speakers had significantly lower centre of gravity measurements than both the rural and urban straight female speakers. Finally, Saigusa (2016) found that the openly lesbian actress Jane Lynch had a lower centre of gravity of /s/ when talking to another openly lesbian journalist than when she spoke with non-lesbian news anchors. Lynch also uses a lower /s/ when discussing LGBTQ topics with the non-lesbian news anchors. These results imply that /s/ may be used to index some level of lesbian identity or a queer identity more broadly.

Studies of fundamental frequency and fricatives show that there is a need for further research in this field. The findings tend to be inconsistent and for some studies the participant pool is quite low. The following section highlights how this current thesis hopes to address some of the gaps in the previous research.

2.6.3 Current Research

As seen by the outline of the research above, there are still many gaps within the research of gay female speech. The work cited above and much of the research on gay female speakers to this point has focused on cis-gendered, white, middle class English speakers either from the United States or Western Europe. While this trend tends to be the norm throughout much linguistic research, there has been a push to be more inclusive and move beyond such a narrow group of participants.

This thesis has worked towards a level of diversity by taking social class into consideration and working with both middle-class and working-class speakers in the UK. Due to the recentness of the work and the age of the participants, the current group of speakers is also of a different generation than those in much of the previous work cited above. As Jones acknowledges, "there is a need, therefore, to consider the impact of current discourses on younger women's identity construction" (Jones, 2018, p. 13). Most of the participants in this thesis are younger than those recruited

for the previous studies and therefore may view sexuality differently from the previous generations, which will add to the continued body of work.

While the participants in this study are still part of the disproportionally large body of work that focuses on Western, white, cis-gendered women, there has been an attempt to broaden who is considered when researching gay female speakers and to include people who have not yet been included. It is hoped that this thesis will add to the body of knowledge that has already been developed, that it will also help to encourage further research on more diverse lesbian speakers and can be used as a starting point for research in the future.

Chapter 3. How Speakers Perform Identity: Production Study

3.1 Introduction

As outlined previously in the introduction, the current research is comprised of two main studies. This chapter covers the first main study, researching the potential connection between speech production and sexual orientation. It details the production study, providing an overview of the methodology, results, and discussion.

The research is based largely on previous sociophonetic work that investigated language, gender, and sexuality. As discussed in the literature review (Section 2.5), fundamental frequency (F0) and fricatives are the most researched sociophonetic variables in association with both gender and sexual orientation. This study focuses on these two elements so as to allow for comparison with previous work.

Initially three main research questions were posed based on the two elements of speech that were investigated, i.e. F0 and the production of /s/. These questions are based on previous findings outlined in the literature review.

1. Is the average F0 of a gay speaker lower than the average F0 of a straight speaker?
2. Do gay speakers have a smaller F0 range than straight speakers?
3. Do gay speakers produce a more retracted /s/ than straight speakers do?

These initial questions guided the research and are the focal point of the main production study. However, in order to take more detailed account of the individual characteristics of the participants, more questions were considered as the investigation continued. As will be detailed below, the data is drawn from speakers who were on a local women's football team and a second set of participants who

were not involved with the team. The group of speakers on the football team predominately identified as gay, while the non-team member speaker group had a higher proportion of straight participants. However, as will be explained in more depth in the methodology (Section 3.2.4), there are gay and straight participants in both team and non-team groups. Along with the research questions presented above, this chapter will also consider the influence team association may have on speech production as well as whether in-group speech is evident.

A second element that will be considered with this data is social class. As will be discussed below, the participants in the study identified their own social class during the course of data collection. Not only did participants identify with different social classes across sexual orientation lines, they did so also across team-affiliation lines. The variation in social class, as well as the differing associations with the football team, allows for a more layered investigation that will consider more than sexual orientation alone. The research reported below will investigate how all these identity factors may interact in speech production. For this reason, a fourth research question was added.

4. How does the interaction of sexual orientation, social class, and team association impact F0 and /s/ production?

It is important to note that due to the sample of participants, statistical models of a three-way interaction between sexual orientation, social class, and team association was not possible. This is addressed further in the methodology section. However, it is still important to consider the possible interactions that can be tested for in the present data.

In connection with the fourth research question, there will also be a presentation of the qualitative data collected from interviews with the participants, predominantly focusing on the gay participants. The views and beliefs expressed by the participants show important variation within and across the different participant groups and present possible explanations for the distribution of phonetic variation among the speakers.

The chapter is presented as follows. First there will be a description of the methodology for the production study. After the methodology, the results will be presented, beginning with qualitative results and then data on fundamental frequency and fricative observations. Following the results, there will be a discussion of what the results signify about speech production among the members of this group of speakers and what it could mean for a larger population.

3.2 Methodology

The inspiration for researching a local women's football team came from knowing one of its founding members. The team will henceforth be referred to as the Yorkshire Town Ladies (YTL). When this project was first conceptualised, the team was experiencing a particularly strong year on and off the field. They had been undefeated for the entire season, only drawing one game, and they were also enjoying spending time together off the pitch as a group of friends. Many of the participants mentioned that specific season as one of their favourite memories while playing for the YTL. However, owing to circumstances beyond the researcher's control it was necessary to alter the study design after the project had started, as the team was no longer dedicated to playing football together as they had been previously.

By the time data collection started, many of the players from that unusually successful season had already left, either through relocating, injury, or simply not having the time to continue playing. The bulk of the recordings from the YTL come from a small, core group who continued to play despite the exodus by other players. However, this core group was substantially smaller than the original team and not large enough to make a full football team for every match. Even before the first recordings took place, there were already several games that had to be forfeited because there were not enough players available. By the point at which the recordings were under way, the team had unofficially disbanded due to lack of participation. Within a few weeks, the team was officially disbanded but they hoped to improve their numbers for the next season. They did not gain more members, and today the team no longer exists.

The project originally sought to follow a community of practice (CofP) approach and so as to understand the speech production within that specific community, with regard to language and sexuality. A CofP is defined by Eckert & McConnell-Ginet (1992, p. 464) as "an aggregate of people who come together around mutual engagement in an endeavour." They go on to clarify that what makes a CofP different from a traditional community of speakers is that it is "defined simultaneously by its membership and by the practice in which that membership engages" (p464). While it is possible that the football team that existed during the season that the project was initially proposed was a strong CofP, it was a much looser community by the time the recordings took place. This change from a strong CofP to a loose CofP was based on the fact that key players were no longer present, there were multiple new players that were not well known to the original team, and, importantly, the majority of members no longer put in consistent effort to participate in. However, due to their association with the team and their existing friendships, they were still a loose community that could be compared to speakers that were not associated with the team at all.

Recordings were collected from all interested members of the YTL. Fortunately, many of the original members are represented in the current data. However, on their own, these recordings are too few to meet the target number aimed for in the design, and therefore additional participants were included. In the initial design of the study, multiple recordings were planned with the YTL in order to obtain data that could be compared according to different contexts (e.g. identity of the interlocutors, conversation themes). With the team disbanding, comparative data was instead gathered through participants that were not associated with the YTL in order to understand how members of the YTL might distinguish themselves through speech.

3.2.1 Pilot Study

A pilot study was undertaken as the researcher established a relationship with the Yorkshire Town Ladies. Participants for the pilot study were recruited through "friend of a friend" associations (Milroy 1980), with the intention of finding speakers who would identify as female, gay, from Yorkshire, between the ages of 25 and 35, and white British. These inclusion criteria were set in order to match the

qualities of the original football team. Four participants meeting these requirements were recruited.

Data was collected through one-on-one interviews recorded in a quiet room using a Zoom H4n recording device, set to capture 16-bit PCM .wav recordings at a sampling rate of 44.1kHz. The locations for interviews were chosen based on neutrality and accessibility, with all recordings being carried out in quiet empty classrooms in university buildings around York. As is common in sociophonetic fieldwork, the recordings did not take place in a recording studio (for recordings conducted outside a recording studio, see Podesva, 2007; Moore and Podesva, 2009; Levon, 2016; Morris, 2017). This method of recording outside of a studio was done to ensure the comfort of the participants, as they might have been uncomfortable in a recording booth in the Language and Linguistic Science Department. Therefore, neutral spaces, like in a university classroom or their own homes, proved to be the best places for recordings, as it kept the recording private and some participants had the comfort of being in their own homes.

At the beginning of the interview, participants completed a questionnaire to establish biosocial data such as age, sexual orientation, city of origin, ethnicity, and level of education (see Appendix 1 for full questionnaire). Questions were specifically created to be open-ended so that participants could identify themselves in any manner they chose, rather than using predetermined categories they might not associate with themselves. All the participants identified as female, gay, and white British, with their ages ranging from 27 to 34, and all were raised in Yorkshire. All four participants had completed a university-level degree and were employed at the time of the interview.

The first part of the recording was a reading task in which participants read a short story aloud (this can be seen in full in Appendix 2). Directly after reading the short story, participants were asked to retell the story to the best of their ability. Retelling the same story ensured that all participants would provide a stretch of semi-spontaneous speech that was on a similar topic and therefore more suitable for comparison. This comparable story was based on Waksler's (2001) study.

After retelling the short story, the rest of the recording consisted of a sociolinguistic interview that focused on experiences and opinions participants had in connection with gender and sexuality, as well as general stories about their childhood and growing up. The interview was conducted in a conversational manner, in the hope that participants would be more comfortable while discussing potentially sensitive topics. This conversational approach is influenced by the idea of *vernacular* influenced by Labov (1966) and was employed to ensure the participants were not overly self-conscious during the interview, particularly when discussing these potentially upsetting topics.

None of the questions was designed to target previous traumatic events; however, due to the sensitive nature of the topics, such as publicly "coming out", it was possible that some of the questions could have triggered some painful memories. I was aware that some of the stories from an individual participant's past could be upsetting if, for example, family members had reacted negatively to her sexual orientation, and I therefore did my best to take cues from the speakers and only pursue topics that they appeared comfortable discussing.

After conducting the four interviews in the pilot study, the recordings were then orthographically transcribed (the interviewer's contributions were included). The sound files were then edited to create separate .wav files during the short story reading, the short story retelling, and the interview for each participant. A duplicate copy of the interview files was created and edited further to remove the interviewer's speech, as well as to remove unusable data, such as when a participant was laughing while speaking, or employing creaky voice. This second edited file was used to collect F0 data, as it contained only the speakers' modal voice. The focus on modal speech was discussed in the literature review above in Section 2.5.1.

Tokens of /s/ were manually marked in Praat (Boersma and Weenink, 2019) based on a combination of auditory and acoustic analysis, relying on the spectrograms and waveforms to mark the beginning and end of the friction. Each token of /s/ was also coded as word-initial, -medial, or -final. Tokens of /s/ that were either directly preceded or directly followed by another sibilant were removed from the corpus, as it was not possible to reliably distinguish the boundaries between the fricatives. An

example of a token of /s/ preceded by another sibilant can be seen in Figure 3. This spectrogram depicts the utterance "songs so" spoken by a pilot study participant. As can be seen, there is not a clear distinction between the [z] in "songs" and the [s] in "so". Due to the lack of clear, consistent boundaries, these tokens were excluded from the analysis.
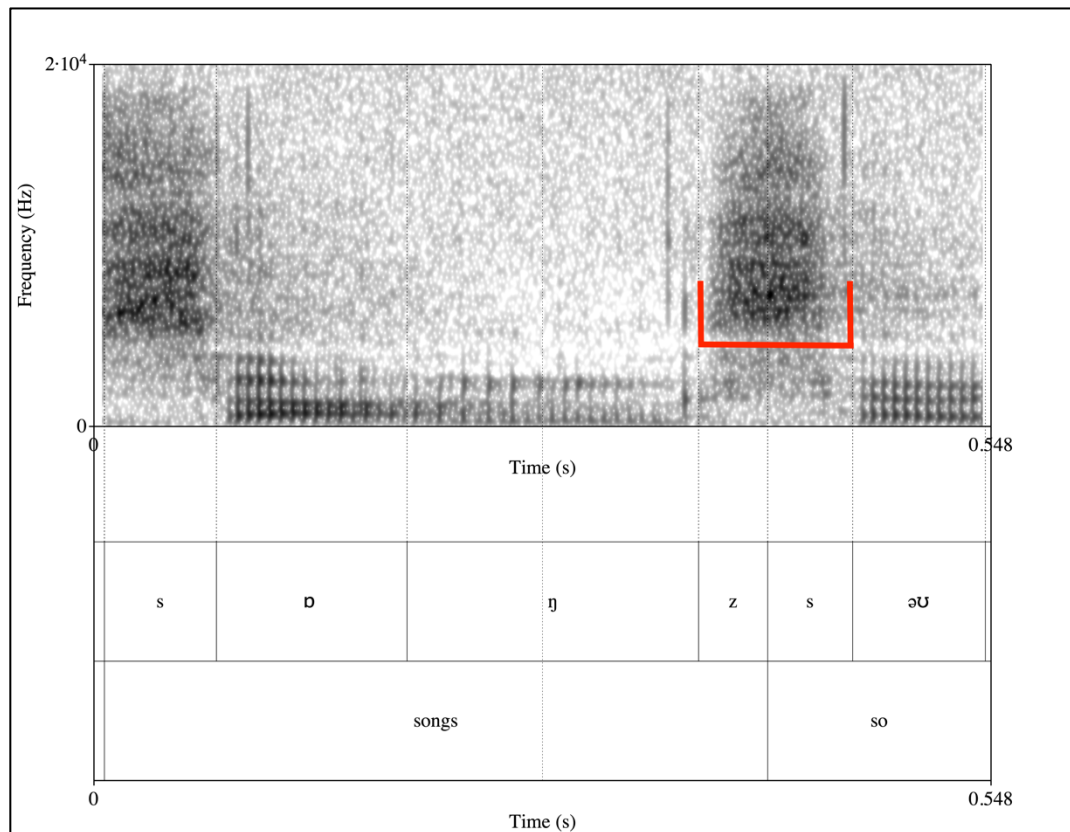
Figure 3 Spectrogram of participant saying "songs so"

Although the participants were asked the same initial interview questions, the interview portions of the recordings were of unequal duration. In order to ensure each participant had a similar number of /s/ tokens to compare, all the interview recordings were analysed for 1020 seconds, centred at the mid-point. This length was determined on the basis of the length of the shortest interview to ensure that all the recordings would be of the same length. The mid-point was chosen because by that point in the interview, speakers would have had time to become accustomed to the recording device. As the reading and retelling portions were all less than 1020 seconds in length, they were each used in their entirety.

For all the spectral measurements, each token was first bandpass filtered to a 1000-22050 Hz bandwidth, following Podesva and Van Hofwegen (2016). This filter was used in order to decrease the influence of ambient sounds coinciding with the fricative by deleting any acoustic data with a frequency of less than 1000Hz. The standard set of four spectral moments (centre of gravity, standard deviation, skewness, and kurtosis) were computed for each token of /s/, as were measurements of duration, amplitude, and peak (amplitude). All of these measurements were extracted using a Praat script (Fecher, 2011).

While the script cited above included multiple /s/ measurements, this thesis only considers centre of gravity measurements. The centre of gravity, or centroid, marks the mean of a spectral curve. Ladefoged (2003, pp. 156–7) describes the centre of gravity as the "equivalent to the point on which a piece of cardboard with the shape of the curve would balance on a pin". Therefore, the centre of gravity can be used as a simple, single measurement to describe tokens of /s/. The centre of gravity is also correlated with how retracted or fronted /s/ is. These different measurements come from a single spectral slice. Centre of gravity has frequently been used as the main /s/ measurement in sexuality studies, included Podesva and Van Hofvegen (2016) and Hazenberg (2012).

While all the research cited above uses centre of gravity as one of the main measurements of /s/s production, there are valid criticisms of the use of the centre of gravity as a key measurement of /s/ in research. As Wrench (1995) illustrates, the sole measurement of centre of gravity does not take into account of the entire shape of the spectrum. The fricative may in fact have more than one peak, which the centre of gravity would not adequately capture. He suggests expanding the measurements collected in order to better describe the entire fricative. This would include measurements to account for the multiple formants which characterise /s/. Although there may be more robust measures when describing /s/, centre of gravity is one of the more commonly used methods and is therefore useful to this study for comparison purposes. For this study, it will be used to match previous research and will therefore allow the data to be more directly comparable.

F0 data was collected using a script in Praat created by Lennes (2016). This script can be found in full at the website in the reference. In a separate study, Lennes *et al.* (2016) applied a bootstrapping procedure in order to determine how many F0 observations would have to be collected so as to have a reliable estimate of a speaker's mean pitch range. It was found that after 34 seconds of analysis, the standard deviation fell below 1 semitone and the data stabilised. The Praat script created by Lennes records a maximum of 720 seconds worth of F0 observations per recording; this provides ample data for analysis, based on their findings.

The F0 Praat script was run on the three sound files created for each participant: reading, retelling, and the edited interview file. For the interviews, the 720-second selection was selected, centring on the mid-point of the recording. This would allow time for speakers to 'warm up' and become accustomed to speaking while being recorded. It also allows for consistency as observations of /s/, discussed above, were also taken from the middle of the recording. Sound files for reading and retelling could be used in their entirety, as none of these files was more than 720 seconds in duration.

The script sampled the audio file every 20 milliseconds, either for its entirety or for the 720-second extract. For each observation, the F0 in that time slot was recorded in Hertz as well as semitones relative to 100 Hz, based on the Lennes *et al.* (2015) study. Both Hertz and semitones were used in the analysis of the data.

Once the data was collected, it was run through the statistical program R (2008). In order to account for pitch observations that were not characteristic of the speaker's modal voice range, observations were excluded that fell below 70 Hz or above 400 Hz. These particular values were chosen for several reasons. First, upon initial analysis of the data, it was found that all participants had a median F0 of 162 Hz or above. Observations below 70 Hz would therefore be unlikely to be part of their modal range and were more likely indicative of creaky voice (Esling and Edmondson, 2010). Similarly, 400 Hz and above would also be outside the modal range and would instead be likely to be typical of falsetto register for this set of speakers. The 70-400Hz range was also chosen to match the study by Waksler (2001), to ensure the data would be comparable across the two studies.

Data for each speaker was drawn from the 5% quantile to the 95% quantile, also based on the procedure followed in the Waksler study. This range was taken separately from each recording for all participants and was chosen in order to account for extreme outliers in the data. These outliers may be due to creaky voice or falsetto that were not eliminated in the 70-400Hz cut-off. The use of the quantile extraction also accounted for two well-known errors in Praat's pitch tracking algorithms: pitch-halving and pitch-doubling (Thomas, 2011). In pitch-halving, two F0 periods are mistaken for a single period. Pitch-doubling is when a half of a pitch period is mistaken for a whole period (Thomas, 2011, p. 37). These automated errors led to inaccurate data that must be removed before conducting statistical analysis.

While data was collected and analysed from this pilot study, the results of the pilot will not be addressed specifically in this thesis. As will be discussed further below, the data from the pilot study was added to the corpus for the main study, and therefore will form part of the larger results section. The purpose of this description of the pilot study was to highlight how data was originally collected and analysed, allowing a trial which guided decision on better options to collect and analyse data in the future.

3.2.2 Yorkshire Town Ladies Recordings

Before recording interviews with the Yorkshire Town Ladies team, small adjustments were made to the questionnaire based on the findings of the pilot study. Two open-ended questions were added asking participants to describe their social class and to specify whether they were regular smokers. Based on further interaction with the YTL, it became apparent that these factors might have a strong impact on the data.

Although social class is regularly considered in sociolinguistic research (Labov, 1966; Trudgill, 1974; Rampton, 2000; Snell, 2010, 2018; Holmes-Elliott and Levon, 2017), initially it was not deemed relevant to the present study; the focus was on a community of practice and sexuality. However, after interacting with the football

team and finding out more about the players, it seemed possible that they would not all identify as being from the same social class. With this indication that social class might vary within the group, it was important to collect social class data to allow for a fuller understanding of how the participants might identify themselves.

A second question was added to find out if participants were regular smokers. This question was based on the findings of Van Borsel, Vandaele, and Corthals (2013), who found that the gay participants in their study were more likely to be smokers than the straight participants. The authors cited this as one potential explanation for the difference in the F0 measurements they recorded. Although the effect of smoking was not further discussed in the Van Borsel *et al.* paper, some of the YTL members were observed smoking and it appeared beneficial to the current study to investigate this further.

Questions were also removed that were deemed unimportant to the research. In the study by Rendall, Vasey, and McKenzie (2008), participants' height and weight were also collected in order to determine if there was any correlation between body type, sexual orientation, and phonetic features. While there were some significant findings in the Rendall *et al.* study, after the pilot study this type of analysis was decided to unnecessarily intrusive to the current study. Questions on height, and in particular weight, were the only questions that were met with any hesitation in participants from the pilot study. It was decided that this was not conducive to the comfortable atmosphere that was being sought for the recording sessions. Therefore, these questions were removed. The updated questionnaire can be seen in Appendix 3.

Along with the questionnaire changes, there was also the addition of a sheet of pictures that summarised the short story that participants were asked to read. Many participants in the pilot study struggled to remember the story in detail and therefore provided fairly short retellings. The pictures were used to help prompt the participants, thereby allowing them to describe the story in more detail so as to obtain longer retelling sections. A similar strategy can be seen in the corpus created by Hellmuth and Almbark (2017). The story stayed the same as the one used in the pilot study, but the images that were supplied can be seen in Appendix 4.

As with the pilot study, the recordings were captured on a Zoom H4n solid-state recorder (sampling rate 44.1 kHz; bit depth 16-bit). All the interviews were carried out as described in the pilot study, with the short story being read and retold, followed by a sociolinguistic interview. However, these interviews tended to be slightly longer, as more questions had been added about football generally and the participants' experiences on the team. A question was also added about the significance religion had in a participant's upbringing. This was based on discussions with a participant in the pilot study who felt religion had a strong connection to her understanding of her own sexual orientation.

The names of all the participants, in order of being interviewed, can be seen in Table 1; all the names are pseudonyms, as are the names used later in this section. With the exception of two of the participants, all the interviews were conducted in the participant's home. Several participants voiced their nervousness about the interview, as they were scared there might be a test or that they could do poorly in some way, despite being assured that this was not the case. As a result, some of the participants were slightly apprehensive at the beginning, particularly about reading aloud. However, by the end of the recording session, all the nervous participants stated that the interview was not as bad as they had feared and that they had actually had fun. This sentiment was very helpful as some did post their positive experience on the football team's social media page, which encouraged more participants to take part.

Yorkshire Town Ladies participants			
1. Michelle	4. Grace	7. Abigail	10. Olivia
2. Natalie	5. Elizabeth	8. Ella	11. Ruby
3. Hannah	6. Taylor	9. Chloe	

Table 1 List of participants from Yorkshire Town Ladies in order of recording

By the end of the six-month collection period, it became clear that no further members would agree to be interviewed. Also, at this point, the team had officially disbanded and most members were not responding on the social media page at all, and therefore could not be contacted for interview requests.

3.2.3 Non-team Recordings

In order to compensate for the disbanding of the Yorkshire Town Ladies team, further recordings were collected outside of the football team. Additional participants were sought that would match the demographic profile of the team, but the search particularly focused on straight women, as nine of the eleven Yorkshire Town Ladies members identified as gay. By having a comparable number of straight female speakers, there would be appropriate data to make valid statistical comparisons between the groups of gay and straight speakers. Via these comparisons it would be possible to obtain solid evidence for or against the existence of a "gay voice" for women.

Through friends of friends, a further 11 participants were recorded. All the new participants identified as White British, female, from Yorkshire, between 27 and 35 years old, and straight. The interviews were conducted in exactly the same manner as those carried out with the Yorkshire Town Ladies, using the same questionnaire and reading tasks. As with the other participants, recording sessions were conducted in their homes or in a neutral area, like a university classroom. Generally, these interviews were shorter than the recordings of the Yorkshire Town Ladies as football was not discussed and the discussions of sexuality did not last as long as they had with Yorkshire Town Ladies team members. The recordings from the pilot study were also added to the larger corpus, as they correspondingly matched the same demographics described above.

The inclusion of the pilot study speakers, as well as the new participants that were not associated with Yorkshire Town Ladies, allowed for comparisons between team members and non-team members, as well as comparisons between speakers on the basis of their sexual orientation and social class.

3.2.4 Full Data collection

With the four participants from the original pilot study, the members of the Yorkshire Town Ladies team, and the additional straight participants, there was a total of twenty-six participants. This amounted to a total of 21 hours and 15 minutes of recordings for all the speakers.

However, some of the sound files had to be edited further or removed from the corpus. Due to technical difficulties, four of the recordings were deemed unsuitable for the study and were removed from the corpus. The interview with Grace was conducted in her kitchen and the recording included the electrical hum from the refrigerator which turned on and off periodically for the entire session. When explored further, it was found that this hum stayed at approximately 100Hz. This created an issue when conducting descriptive statistical work on Grace's F0 measurements, as there were a large number of F0 measurements from the hum of the refrigerator in what had otherwise been silent pauses between speech. Using the software Audacity (2019), a notch filter with a centre frequency of 100Hz was implemented. As Grace's modal F0 was regularly observed far above 100Hz, this extraction should not have had a significant impact on her F0 measurements. It also did not have an impact on the /s/ data collection, as the frequency band below 1kHz was removed all together before collecting any measurements of /s/. The application of this filter was not judged to affect the suitability of the sound file for further analysis.

With four files removed, the corpus consists of 22 participants and stands at 18 hours and 10 minutes of recordings. A full list of participants, their connection to the football team, and their sexual orientation can be seen below in Table 2. The rest of the methodological description that follows is only applicable to these participants.

Yorkshire Town Ladies		Non-football	
Gay	Straight	Gay	Straight
Michelle	Taylor	Heather	Scarlett
Natalie		Laura	Maria
Grace		Sarah	Julia
Elizabeth		Rebecca	Isabella
Abigail			Lucy
Ella			Ava
Olivia			Lily
Ruby			Emily
			Jessica

Table 2 Full list of participants in study

3.2.4.1 Participant profile

Of the 22 participants, 12 self-identified as gay based on the questionnaires they were asked to fill out at the beginning of the interviews, while the remaining 10 identified as straight. The majority of the participants that identified as gay were also part of the Yorkshire Town Ladies football team. However, there is one participant from the team that identified as straight, and four self-identifying gay participants from the pilot study who were not part of the team. This allows for some consideration across sexual orientation and participation in the football team, but not strong statistical testing.

The age span was from 21 to 46 at the time of recording. With the exception of Grace (the youngest participant) and Elizabeth (the oldest), participants were aged between 27 and 35 at the time of recording. Only 3 participants identified themselves as regular smokers, so the possible effects of this factor were not investigated any further, as there was not enough data. Despite one of the pilot study participants indicating that religion played a significant role in her childhood, as well as impacting her understanding of her own sexual orientation, the vast majority of the participants stated that they were not raised religious and that religion did not have a large impact on their life. For this reason, religion was not considered further.

Participants were asked to state their social class on the questionnaire. As gender and sexual orientation were both self-identified by the participants, social class was treated in the same way. The social class categories presented for each speaker are based on how individuals identified themselves, and not on factors such as career, education or income. Though there may be inconsistencies in categorisation because participants understand their own social class differently, I argue that continued use of self-identification contributes important information on how the participants view themselves. Chun (2019, p. 342) writes, "it is important for sociolinguists to find ways to present people's lived realities and identities of class in all their complexities." While this singular question may not have considered all the complexities of class, it did take into account "people's lived realities".

Some participants wanted to discuss this social class question further, as they were not sure how to answer. In these cases, when they were filling in the questionnaire and got to the social class question, they might have said something similar to "This is a tricky question." In these situations, I would ask the participants if they would be comfortable talking about it while being recorded, and they all agreed to this. During the interview, I would then ask questions about their views on social class in the UK and what constituted certain social classes based on their experience – for example "what makes a person middle-class?" By the end of our discussion, those that had expressed hesitation or confusion all had an answer that they would then use in the questionnaire.

However, one participant, Ruby, labelled herself as both working- and middle-class on the questionnaire, and therefore could not be classified as just one or the other. The pilot study did not ask did not ask participants to specify their social class; however, one participant spontaneously identified her social class in our interview. In the group, 10 participants identified as working class, 8 identified as middle-class, 1 as working-class/middle-class, and 3 did not respond. Table 3 below shows how each participant was coded for sexuality, social class, and team membership.

Participant	Sexuality	Social class	Team member
Abigail	gay	middle	yes
Ava	straight	working	no
Ella	gay	working	yes
Elizabeth	gay	working	yes
Emily	striaght	middle	no
Grace	gay	working	yes
Heather	gay	unknown	no
Isabella	straight	working	no
Jessica	straight	middle	no
Julia	straight	middle	no
Laura	gay	unknown	no
Lily	straight	middle	no
Lucy	straight	middle	no
Maria	straight	working	no
Michelle	gay	middle	yes
Natalie	gay	middle	yes
Olivia	gay	working	yes
Rebecca	gay	working	no
Ruby	gay	unknown	yes
Sarah	gay	unknown	no
Scarlett	straight	working	no
Taylor	straight	working	yes

Table 3 Participants with sexuality, social class, and team membership

As can be seen in Table 3 and as stated above, the vast majority of the YTL are gay and only one team member is straight. This means that it is not possible to fit a two-way interaction between sexual orientation and team membership. However, it is possible to consider interaction between sexual orientation and social class, as well as team membership and social class.

3.2.4.2 Creating Data Set

As with the pilot study, the recordings were divided into the reading section, retelling section, and interview. A fourth sound file was created from which laughter, excessive creaky voice, or falsetto had been removed. Upon reflecting on the successes and difficulties of the pilot study, it was decided that it would be best to use all recordings in full, regardless of length, so as to maximise the data. Instead of working from the file's temporal midpoint to collect F0 and /s/ observations, as

had been done in the pilot study, tokens were collected from the entire duration of the recordings for the main study.

The sound files for reading, retelling, and the edited interview files for each participant were also run through the Lennes (2016) Praat script to collect the F0 data for the participants. While it was stated above that Lennes *et al.* (2016) found that one only needed to consider a brief amount of voiced speech in order to accurately determine mean F0 measures, it was decided that it would be beneficial to obtain F0 measurements for the entire duration of the interview. This method would give a full sampling for every participant, which removed the concern about having to deduce the best point in the sample at which to begin collecting data.

The F0 data was edited to remove outliers, non-modal speech, and to account for pitch doubling and halving, as discussed above. F0 values below 90Hz and above 400Hz were excluded. As discussed above, speakers categorised as female in previous work have typically had an F0 mean around 200Hz. Since all the participants in this study identify themselves as female, 90Hz was deemed an appropriate cut off point for the bottom of their modal range and 400Hz was an acceptable cut off point for the top of their modal range. While these ranges may not account for all speakers that identify as female, initial scans of the data indicated this range would be appropriate for the present group of participants.

The minimum of 90Hz was increased from the 70Hz threshold used in the pilot study to further account for creaky voice, as this feature was prominent in this group of speakers. Data were removed that were two or more standard deviations above or below the mean, on an individual speaker basis. The change to remove outliers according to standard deviation instead of quantiles was also included to account for creaky voice in the recordings. Due to creaky voice being frequently present in the data, basing the bottom of the threshold on the 5% quantile did not consistently eliminate evidence of creak. However, basing the threshold on individual standard deviations did eliminate creak data.

Each sound file was transcribed using the software ELAN (2018) in order to time-align the signal with its corresponding transcription. After the transcription was

complete, the files were forced-aligned using the software package FAVE (Rosenfelder *et al.*, 2014). This produced a Praat TextGrid aligning the transcribed phonemes with the relevant sections of the audio file.

All files were then manually checked in order to ensure the /s/ tokens were correctly identified and the boundaries precise. Observations of /s/ were collected for the entire duration of the recording, as with the F0 observations. As before, all /s/ tokens that were adjacent to another sibilant were removed (Figure 3). Tokens were also removed if there was interference from ambient noise or overlapping speech. Finally, /s/ tokens in which voicing was present were removed. The remaining tokens were run through a Praat script, which would log each occurrence of /s/ for centre of gravity, standard deviation, skewness, kurtosis, duration, amplitude, and peak. This script adapted from Fecher (2011) can be seen in Appendix 5.

Each /s/ token was coded for syllable position (onset, coda, and ambisyllabic), phrasal position (phrase-initial, -medial, -final), and prosodic prominence (stressed vs. unstressed). Having consulted previous work, including Podesva and Van Hofwegen (2016) and Stuart-Smith *et al.* (2003), syllable position was prioritised in this study, versus the position in the word as per the pilot study design. For a similar reason, phrase position was also labelled.

All /s/ tokens were coded for the syllable in which they appeared. However, it was not possible to determine whether some /s/ tokens were part of the coda of one syllable or in the onset of the next. /s/ tokens in cases of this type were labelled *ambisyllabic*. Predominantly, these ambisyllabic tokens appeared in intervocalic consonant clusters in words such as "extra".

Coding for placement in the phrase was based on Stuart-Smith, Sonderegger, Rathcke, and Macdonald's (2015) study of stop consonants. A phrase was defined as "the interval between two intervals of silence of at least 150 ms" (Stuart-Smith *et al.*, 2015, p. 515). If the /s/ token was in the first word of a phrase it was labelled as phrase-initial, and if it was in the last word of a phrase it was labelled as phrase-final. All other tokens were labelled as phrase-medial. An example of syllable and phrase coding can be seen in Figure 4 below.
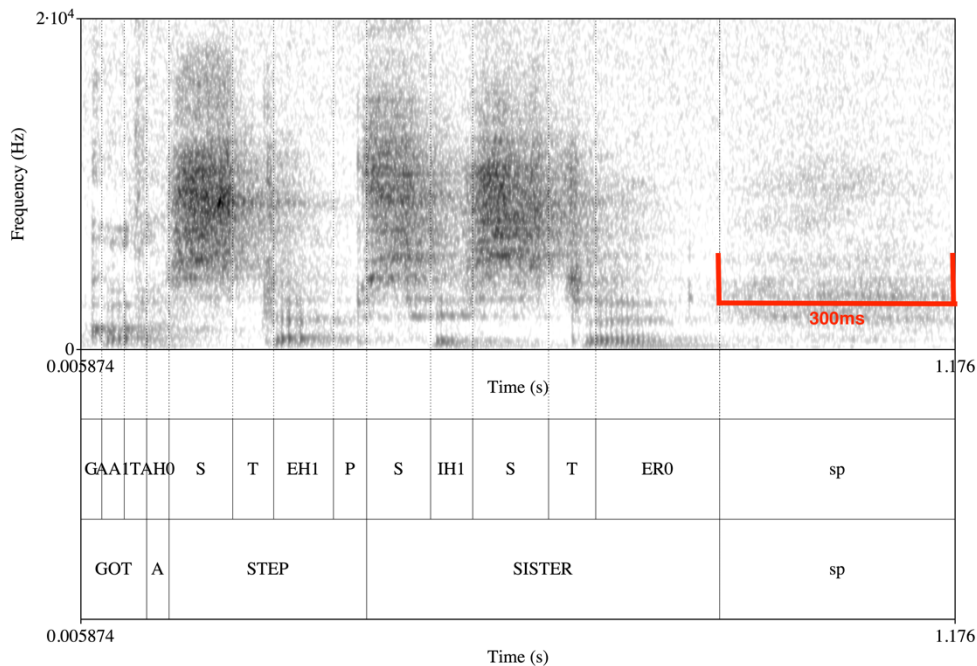
Figure 4 Spectrogram of Natalie's production of "got a step sister", followed by a 300ms pause

In Figure 4, there is speech and then a pause, labelled as "sp", which exceeds 150ms. In this case, the /s/ token in the word "step" is classified as phrase-medial, but the two /s/ tokens in the word "sister" are both classed as phrase-final. The /s/ token in "step" and the first /s/ token in "sister" are both classed as syllable-onset, while the second /s/ token in the word "sister" is coded as syllable-coda.

Finally, syllable stress was considered. In previous research, such as Holmes-Elliott & Levon (Holmes-Elliott and Levon, 2017) or Podesva and Van Hofwegen (2016), /s/ tokens have been described as stressed or unstressed. However, in practice making this distinction was not straightforward. As Cruttenden (1986) acknowledges when discussing stress, "[a]ny description of English word-stress rules inevitably involves a large number of exceptions" (p. 19). In order to match previous research like Holmes-Elliott & Levon (2017) and Podesva and Van Hofwegen (2016), and to simplify stress classification, syllable stress was treated as a simple binary distinction *prominent* versus *not prominent*. Therefore, even when an /s/ token appeared in a word with multiple stresses, all levels of stress were simply counted as stressed without separating the levels of stress. If the syllable was not judged to be stress-bearing at all, it was marked as unstressed, or not prominent.

In the example in Figure 4 above, both /s/ tokens in the word "sister" are in a prominent syllable, and therefore both are coded as stressed.

In order to ensure best practice when classifying the stress patterns, a subset of the material was independently coded for syllable stress by one of the project's supervisors, using the same criteria. Olivia's interview was chosen, as it is one of the longer interviews and makes up roughly 5% of the total corpus. An agreement rate threshold of 80% was deemed satisfactory for present purposes, bearing in mind that judgements of relative prosodic prominence are inevitably subjective (Wells, 2006, p. 248).

The final corpus, at 18 hours 10 minutes in length, yielded 12,368 /s/ tokens and 668,786 F0 observations across all 22 participants.

3.3 Results

This section will outline the results based for the 22 speakers in this study. Qualitative results based on the interviews with the participants are presented first. These results will include the participant's views on how they labelled their sexual orientation, whether participants describe themselves as activists, and if there is a "gay voice" for women. The first two sections of the qualitative data will focus on the gay speakers, but the third section, on the concept of a "gay voice", will include data for all the participants. These results will give a more detailed view of the individuals that took part in this study, as well as providing important laypeople's insights into the idea of a gay voice.

Secondly, the quantitative data will be presented. This will begin with data on fundamental frequency (F0) means and other statistics relating to speaker pitch, including F0 range. Finally, data from the /s/ productions will be described. For all three sections, the data will be presented as follows: first, across sexual orientation groups, second, considering self-identified social class, and, third, according to association with the Yorkshire Town Ladies team.

3.3.1 Qualitative results

In this section, a summary of some of the key questions from the interviews will be presented. Naturally, there was variation within the groups on their views of topics like LGBTQ activism and the existence of a gay voice, but there were also interesting similarities between the speakers that are worth investigating further in this study. These qualitative results give a deeper understanding of the participants as individuals, and allow us to see the diversity of self-identification, even if they are part of the same sexual orientation, class, or team category.

To begin this section, there will be a detailed description of how participants felt about labels for sexual orientation preference and what terms they preferred themselves. Second, I will consider the importance of activism among the current group of speakers. The last section of the qualitative results will present the participants' views on the idea of a "gay voice" and whether they believe such a thing exists, and what it might sound like.

3.3.1.1 Preferred terminology

As mentioned previously, many of the participants expressed a preference for the term "gay" and said that is how they would describe themselves, which is why this thesis has chosen to use the term throughout. This section will outline how participants described their sexual orientation and the reasons for their preferences.

In the questionnaire, participants were asked, "How would you describe your sexual orientation?" and were given free space to write in what they preferred. Of the 12 gay participants, half wrote "lesbian" and half "gay". Interestingly, only two members of the Yorkshire Town Ladies team, Grace and Michelle, described themselves as "lesbian" on the questionnaires, while the rest of the team described themselves as "gay". When asked in the interview if this was the term she preferred, Grace expressed how much she disliked the word "lesbian" and said, "I hate that word. I absolutely hate that word." Other participants shared these feelings towards the word "lesbian". Heather said that while she did not mind the word any longer, she used to hate the word "lesbian" and said, "Yeah I thought it sounded like a disease or something". Ruby also expressed a disliking for the word "lesbian" and

describes it as "terrible". While others in the study did not express as strong a dislike for the word, not a single participant preferred it to other terms.

Sarah and Rebecca, who both answered "lesbian" on the questionnaire but said they preferred "gay", described why they think "lesbian" is a less positive description of their sexual orientation. Sarah commented that she preferred the word "gay" because "it's not a noun". She acknowledged that it would be equally inappropriate to call a group of gay men "gays" and therefore using nouns to categorise people was not something she would feel comfortable about doing. Rebecca echoed this sentiment, saying that she did not like "being **a** lesbian" (emphasis noted in her voice). Both speakers appear to prefer the term "gay" as it is an adjective they said that simply describes part of who they are, instead of a noun that puts them entirely into a single category.

Ella described herself as "gay" on the questionnaire and when asked if that was her preference, she said she was not bothered by any terminology, but did prefer "gay". She reflected, "Like if someone says you're a lesbian it's like you're a one to yourself and more boxed but when I'm gay I'm with everybody else." Natalie also commented that lesbian is "just not a nice word to say." While six participants say that they did not mind what they were called, as long as it is not said in a hurtful way, three do clarify that they still prefer gay and that is how they would be likely to describe themselves.

These views expressed by the participants demonstrate an interesting dichotomy between the formal terms one uses when filling in surveys and forms collecting biosocial data, and the terms one prefers. Of the six participants who wrote the term "lesbian" on their questionnaire, four of them said they prefer the term "gay" and the other two simply said they did not care about terminology at all. A fuller consideration of terminology, and how it fits within current social movements, will be addressed further in the discussion in Section 3.4.1.

3.3.1.2 Connections to activism

In order to find out how they may connect with a wider LGBTQ community and how they might view the broader significance of being a gay person, participants were asked if they considered themselves activists. Owing to the conversational nature of the interview, not every gay participant was asked about activism, but the majority were.

Only two participants, Heather and Rebecca, said that they did identify on some level as LGBTQ rights activists, and Laura acknowledged that she would put in effort to support friends who are activists or fighting for rights, but was hesitant to call herself an activist. Interestingly, all three of these participants were in the pilot study sample; none of the members of the Yorkshire Town Ladies said they were activists.

The view shared by Abigail, Elizabeth, Grace, Natalie, and Ruby, all the members of the Yorkshire Town Ladies, was that they supported the movement for equal rights and saw it as valid, but would not actively participate themselves. Elizabeth summarised the views of the majority of the team when she said, "Would I stand on a front line and wave a banner? I'd probably not unless it directly affected me". Ruby acknowledges that she would rather spend time on more enjoyable activities and Grace says that the time spent working towards activism would be wasted, as mind-sets were unlikely to be changed by these endeavours.

When asked about LGBTQ activism, Natalie said, "I just just get on with my life". Her viewpoint was shared by many of the other YTL members. This echoes similar feelings acknowledged in the previous section about terminology and labelling, and the preference for the term "gay" because it is an adjective that only described part of a person, opposed to "lesbian", which was too strict a categorisation for many.

Through the discussion of activism and preferred terminology, it appears that the gay YTL members do not foreground their sexual identity over other possible identity markers. Sexual orientation might very well be relevant in certain contexts and in particular situations; however, it may not be an important identity on a day-

to-day basis for these speakers. This possibility will be considered further in the discussion section.

3.3.1.3 Views on the existence of a "gay voice"

In the interviews, participants were asked if they believed a "gay voice for women" existed. Many were unsure and a few referenced physical appearance as being a more telling factor than voice when guessing a woman's sexual orientation. Multiple participants said the question itself was interesting and something they had never thought of before being asked. Many were aware of the idea of a male "gay voice" and discussed the qualities of that voice, but said, until being prompted, they had never thought about an equivalent for women. After considering it, participants varied in respect of whether they believed it existed. Some did not believe a "gay voice" for women existed because they had never noticed it. Others said that while they had never considered it before, its existence was plausible; as Rebecca put it, "there must be something 'cause I have quite good gaydar". "Gaydar", defined by Sulpizio *et al.* (2019, p. 1), is "the ability to recognize sexual orientation" in other people without being directly told.

Half of the participants said they did not believe a gay-sounding voice existed for women. However, of those 11 participants, three went on to say that while they did not believe there was a gay-sounding voice, there might be other factors that would hint at a woman's sexual orientation, such as style of dress or use of mannerisms. Another five of these participants who disputed the existence of a "gay voice" clarified that there might be linguistic cues associated with women who are "stereotypically gay" or "butch", these terms being used by the speakers. Thus, the participants indicated that a "gay voice" might exist, though it is only employed by women who present a certain type of gay identity.

However, among those who accepted the possibility that a gay voice might exist, a common theme emerged regarding the supposed quality of the voice. Four of the participants mentioned that a "gay voice" for women might sound more masculine, or users of this voice quality sounding like "lads". For example, Michelle says, "But I could probably see how you could base it on how somebody talks 'cause if they're

all bit more masculine, deeper voiced and and then that going with the dress style…". Three of those who did not directly relate a "gay voice" for women to masculine features described the voice as being "rough", either through excessive swearing or through associations with regional accents.

While most of the language about a "gay voice" for women was quite vague, there was one quality that was described as potentially being part of a "gay voice" for women: low pitch. As seen in the quote from Michelle above, a "gay voice" could be associated with a "deeper voice". Other terms used to describe this quality were "baritone" by Ava, "lower tone" by Jessica, "lower voice" by Maria, and "huskier voice" by Olivia. This was the only specific language any of the participants used to describe the quality of a "gay voice" for women.

The following section will look more closely at the speech production of all the participants, including a comparison of pitch according to the speakers' professed sexual orientation preferences. However, perceptions of a "gay voice" will be considered in much more detail in the perception study in Chapter 4, and these perceptions by listeners are specifically focused on mean pitch.

3.3.2 Fundamental Frequency Results

This section will now present the quantitative results. The distribution of the F0 observations for each speaker can be seen in Figure 5 below. Each speaker is labelled under her respective F0 distribution. Wider parts of the plots indicate a greater number of observations. The black diamond for each speaker marks the mean F0 for all of her data.
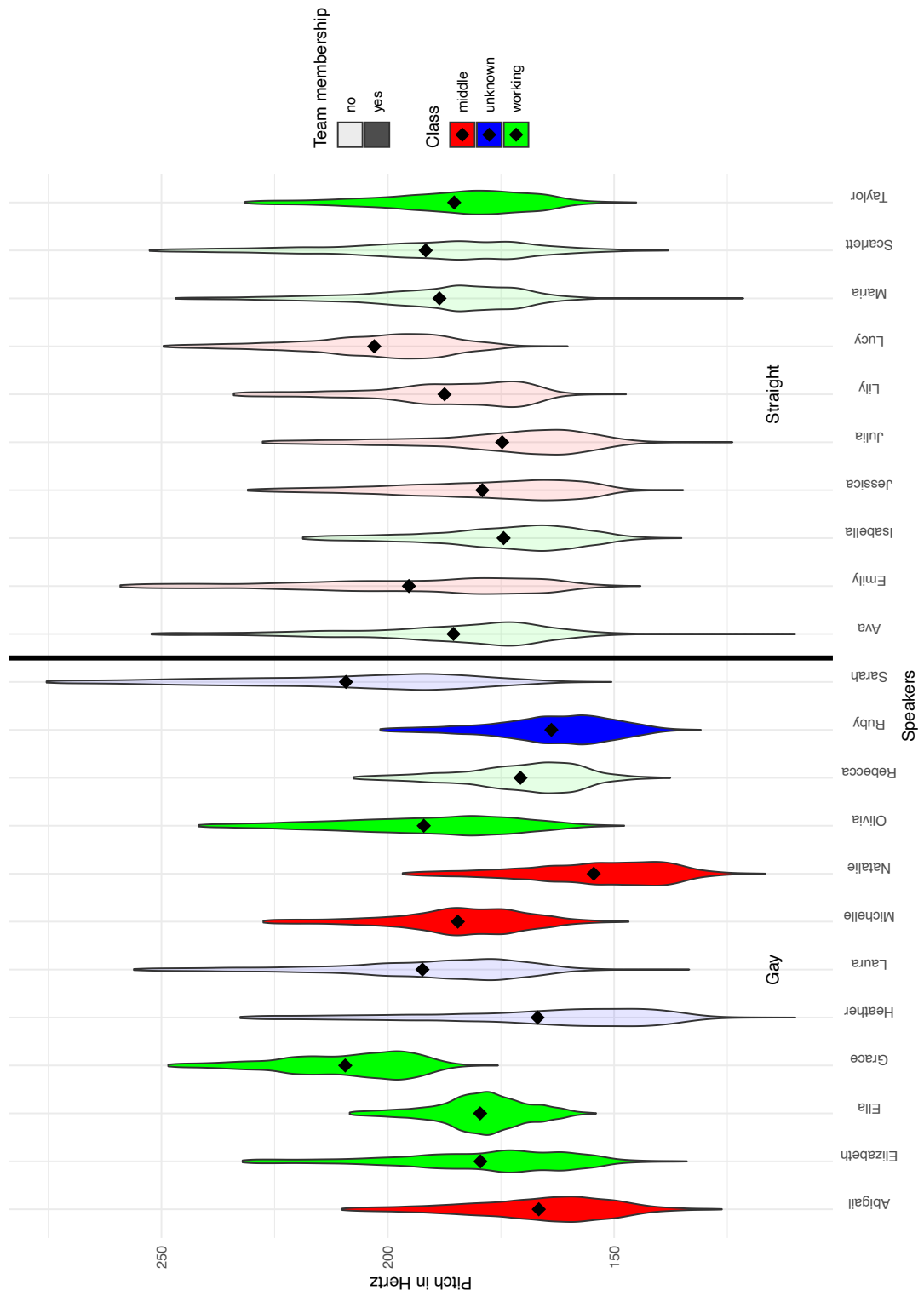
Figure 5 F0 observations for each speaker

3.3.2.1 Mean Fundamental Frequency

As can be seen above in Figure 5, the speakers with the lowest mean F0 measurements identify as gay. This is confirmed by the two groups' pooled figures, as shown in Table 4.

Sexual orientation	Mean F0 in Hertz
Gay	181
Straight	186.5

Table 4 Mean F0 for the gay and straight speaker groups

Table 5 below provides the mean F0 measurements for all the speakers across all three sections of the recordings. They are divided by sexual orientation and then ordered from lowest mean to highest.

Gay Speakers		Straight Speakers	
Speaker	Mean F0 (in Hertz)	Speaker	Mean F0 (in Hertz)
Natalie	154.5Hz	Isabella	174.4Hz
Ruby	163.9Hz	Julia	174.7Hz
Abigail	166.6Hz	Jessica	179.1Hz
Heather	166.9Hz	Taylor	185.3Hz
Rebecca	170.7Hz	Ava	185.5Hz
Elizabeth	179.5Hz	Lily	187.5Hz
Ella	179.6Hz	Maria	188.6Hz
Michelle	184.5Hz	Scarlett	191.6Hz
Olivia	192Hz	Emily	195.3Hz
Laura	192.3Hz	Lucy	203Hz
Sarah	209.2Hz		
Grace	209.4Hz		

Table 5 Mean F0 for all speakers, in order of F0, divided by sexual orientation

In Table 5, it is seen that the speakers with the five lowest mean F0 measurements are all gay. However, three of the speakers with the five highest F0 measurements also identify as gay. The mean values of the gay speakers span a broader range than that of the straight speakers. Yet in spite of this broad range, it is clear that the gay speakers on the whole have lower mean F0 values than those of the straight speakers.

Mixed-effects regression modelling in R (2008) were run using the "lme4" package and the "lmerTest" package was used to summarise the results and produce a *p*-value. As the majority of the participants that were on the Yorkshire Town Ladies

also identified as gay, these two factors cannot be compared together. The social factors are therefore first presented separately, to see how each individual factor interacted with F0 measurements and then there will be model which considers the interaction between social class and team association.

The first regression reveals that the gay and straight speakers are significantly different in respect of F0. The results of the test can be seen in Table 6 below.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	180.9523	0.0339	5337.45	0.000
Sexual orientation (straight)	5.5584	0.0609	91.27	0.000

Table 6 Sexual orientation regression model

There is also a difference according to social class. Middle-class speakers are likely to have lower F0 measurements than those who identify as working-class. Since the undetermined class participants do not make a unified group, they were not included in both the summary statistics and the mixed effects regression models. The mean F0 measurements for social class can be seen in Table 7 below.

Social Class	Mean F0 in Hertz
Middle Class	178.1
Working Class	185.2

Table 7 Mean F0 according to social class

This difference is statistically significant, and the results of the regression model can be seen in Table 8 below.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	178.14395	0.04909	3628.6	0.000
Social class (working)	7.09751	0.0608	116.7	0.000

Table 8 Social class regression model

As with social class and sexual orientation, association with the team is shown to have a significant effect upon the F0 observations for these participants. Those who were part of the Yorkshire Town Ladies have lower F0 means than those who were not on the team. These mean F0 results can be seen in Table 9 below.

Yorkshire Town Ladies	Mean F0 in Hertz
On team	179.7
Not on team	186.7

Table 9 Mean F0 according to team association

This difference in team association is also statistically significant, and the result of the regression model can be seen in Table 10 below.

Fixed effects	Estimate	Std. error	*t*	p-value
(Intercept)	186.67604	0.04276	4365.7	0.000
Team association (yes)	-7.01141	0.0566	-123.9	0.000

Table 10 Team association regression model

Importantly, the categories do not exist separately, and they inherently interact with each other. Ideally, regression models would be run in order to understand the F0 observations when taking in to account all three identity factors. However, due to the confounding participants that identify as gay and are on the YTL, this was not possible. For this reason, regression models were run to consider social class and team association, as team association demonstrated a slightly larger difference in F0 realisations. The results of this regression test can be seen in Table 11. In order to account for all the team data, the undetermined social class speakers were included. Speaker was also treated as a random factor in order to accommodate individual variation.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	187.921	5.966	31.497	0.000
Social class (unknown)	1.569	9.743	0.161	0.8741
Social class (working)	-5.764	9.438	0.683	0.5043
Team association (yes)	-19.36	9.743	-1.987	0.0643
Class (unknown): team (yes)	-6.272	18.277	-0.344	0.7352
Class (working): team (yes)	26.381	12.889	2.047	0.0574

Table 11 Pitch regression model according to two identity factors

When considering the factors together, team association and social class do not appear significant. These associations and multiple identity factors will be investigated further in Section 3.5.

3.3.2.2 Fundamental Frequency Range

Having considered the mean F0 measurements, as well as the patterns in the raw F0 data, it is important to consider the F0 range speakers used. While in the section above the data were presented in Hertz, the following results for F0 range will be presented in semitones. For a discussion on the use of semitones with regards to range, see Section 2.5.1.

Table 12 presents the average range for speakers according to sexual orientation.

Sexual orientation	Range in semitones
Gay	8.672
Straight	9.923

Table 12 Mean range for participants according to sexual orientation

In this table, it is possible to see that the gay speakers have a slightly narrower pitch range than that of the straight speakers based on the entire recording. In Table 13 below, the pitch ranges for all speakers are presented, categorised by sexual orientation and ordered from lowest to highest.

Gay Speakers		Straight Speakers	
Speaker	F0 Range (in semitones)	Speaker	F0 Range (in semitones)
Ella	5.2	Lucy	7.7
Grace	6	Lily	8
Rebecca	7.1	Taylor	8.1
Ruby	7.5	Isabella	8.3
Michelle	7.6	Jessica	9.3
Olivia	8.5	Emily	10.1
Abigail	8.8	Scarlett	10.5
Natalie	9.1	Julia	10.5
Elizabeth	9.5	Maria	12.3
Sarah	10.5	Ava	14.4
Laura	11.3		
Heather	13		

Table 13 Mean F0 range for all participants in semitones

Table 13 shows that there is considerable variation between speakers, even those within the same sexual orientation groups.

As with the F0 regression models presented above, the F0 range models presented are all mixed effects regression models, created using the "lme4" package and summarised using the "lmerTest" package.

The F0 range between speakers according to sexual orientation is not statistically significant; the results of the regression model can be seen in Table 14.

Fixed effects	Estimate	Std. error	*t*	p-value
(Intercept)	8.6722	0.6267	13.839	0.000
Sexual orientation (straight)	1.2512	0.9295	1.346	0.193

Table 14 Range regression model according to sexual orientation

The mean F0 ranges according to social class can be seen in Table 15 below.

Social Class	Range in semitones
Middle Class	8.89
Working Class	8.99

Table 15 Mean range for participants according to social class

F0 range is also not significantly different between social class identities. The results can be seen in Table 16.

Fixed effects	Estimate	Std. error	*t*	p-value
(Intercept)	8.89462	0.78357	11.351	0.000
Social class (working)	0.09966	1.05712	0.095	0.926

Table 16 Range regression model according to social class

However, there are differences in F0 range according to team association. The mean F0 range totals are presented in Table 17.

Team Association	Range in semitones
On team	7.814
Not on team	10.229

Table 17 Mean range for participants according to team association

The effect of team association emerges as significant, as can be seen in Table 18.

70

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	10.2288	0.5254	19.47	0.000
Team Association (yes)	-2.4149	0.8215	-2.94	0.0081

Table 18 Range regression model according to team association

As there is only one range measurement per speaker, it is not possible to run a mixed-effect regression model with all three identity factors considered, due to low quantities of data. However, consideration of how these three factors may interact will be included in the discussion in Section 3.5.

In order to directly compare the ranges of each speaker, data from just the story retelling section will be considered. The retelling portion can give particular insight into range comparisons because they share a similar context. For example, one speaker during the interview may have been particularly excited talking about a subject that another speaker was discussing in a neutral way, and therefore did not use her full modal range. If this is the case, the comparison would fail to reflect a representative picture of the speakers' typical F0 ranges. In the retelling data, participants were all sharing the same information in the same context and therefore their data has the advantage of being more directly comparable. Table 19 presents the mean range for speakers according to sexual orientation.

Sexual orientation	Range in semitones, retelling portion
Gay	7.9
Straight	8.8

Table 19 Mean range for participants according to sexual orientation with retelling data

As with the F0 ranges derived from each speaker's entire interview, which were presented above, the gay speakers have a slightly lower range than the straight speakers. Table 20, below, presents the ranges for all of the speakers, based only on their retelling portion of the recordings.

Gay Speakers		Straight Speakers	
Speaker	F0 Range Retelling (in semitones)	Speaker	F0 Range Retelling (in semitones)
Ella	4.9	Lily	5.8
Grace	5.3	Lucy	6.3
Rebecca	6.1	Taylor	6.8
Ruby	7.3	Isabella	7.6
Michelle	7.4	Maria	8.1
Abigail	7.6	Jessica	9.2
Olivia	7.8	Emily	9.6
Natalie	8.8	Scarlett	10
Elizabeth	9.4	Julia	10
Heather	9.72	Ava	14.3
Sarah	9.8		
Laura	10.5		

Table 20 Mean F0 range (semitones) for all participants, story retelling data

As with the full data results, the difference between gay speakers and the straight speakers is not significant. For the results of the regression model, see Table 21 below.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	7.8871	0.6172	12.78	0.000
Sexual orientation (straight)	0.8884	0.9155	0.97	0.343

Table 21 F0 range regression model according to sexual orientation, story retelling data

Table 22 presents the F0 range from the retelling portion according to social class.

Social Class	Range in semitones
Middle Class	8.09
Working Class	8.03

Table 22 Mean range for participants according to social class with retelling data

F0 range according to the retelling portion is also not significantly different between social class identities. These results can be seen below in Table 23.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	8.08737	0.81346	9.942	0.000
Social class (working)	-0.05081	1.09138	-0.047	0.963

Table 23 Range regression model according to social class

When considering only the story retelling data, there is not a significant difference in F0 range across team association. The mean F0 range totals are presented in Table 24 and the regression model in Table 25.

Team Association	Range in semitones
On team	7.26
Not on team	9

Table 24 Mean range for participants according to team association

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	9.0078	0.5526	16.3	0.000
Team Association (yes)	-1.7522	0.864	-2.028	0.0561

Table 25 Range regression model according to team association

A significant difference in pitch range across team association is only evident in the entire recording data.

Finally, tests were run in order to see if there was any correlation between mean F0 and F0 range. First the data was tested for normal distribution. Using the Shapiro-Wilks test, it was found that the ratings for F0 mean and F0 range were not significantly different from normal: F0 mean (W = 0.98042, p = 0.9225) and F0 range (W = 0.97604, p = 0.8443). Speaker mean F0 was found not to be correlated with speaker F0 range (Pearson's r (20) = -0.016 p = 0.9436).

3.3.3 Results for /s/

This section will focus on the results of the /s/ analysis. As outlined in the literature review, Section 2.5.2.1, the majority of language and sexuality research describing /s/ production has primarily relied on centre of gravity (CoG) as the main measurement. For this reason, this section will mainly focus on CoG in describing the /s/ data.

The CoG measurements for each speaker can be seen in Figure 6 below. This figure summarises all the /s/ observations for each speaker, though the observations will be teased apart according to factors such as surrounding sounds, placement, and stress.
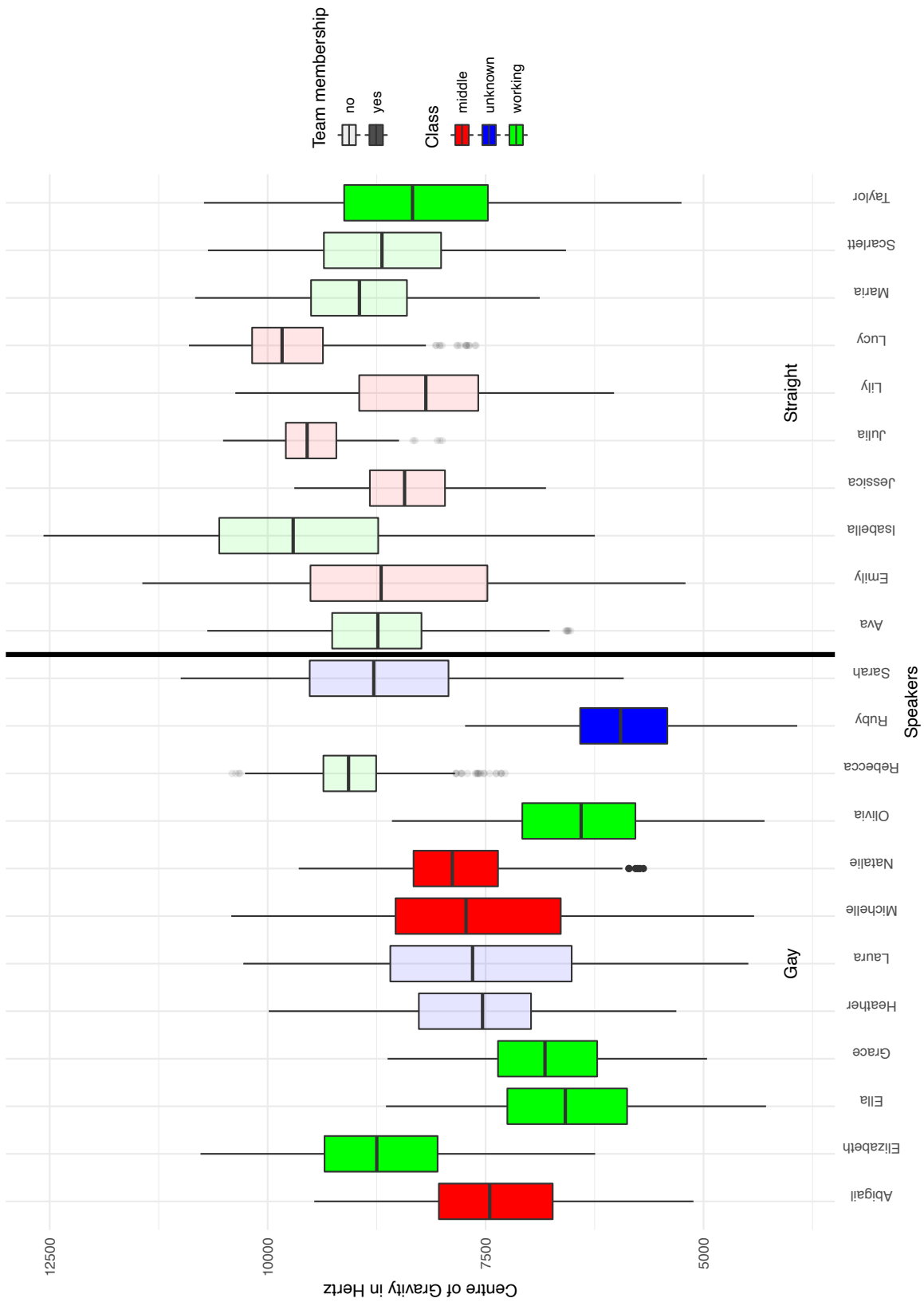
Figure 6 CoG observations for each speaker

As can be seen in Figure 6, gay speakers tend to have a lower CoG in /s/ realisations. This pattern is maintained when factoring out variation in the surrounding context. In order to demonstrate the consistency of lower CoG for the gay speakers, a specific context of /s/ observations was selected. Figure 7 shows the CoG observations for stressed syllable-onset /s/ tokens that occurred phrase-medially.
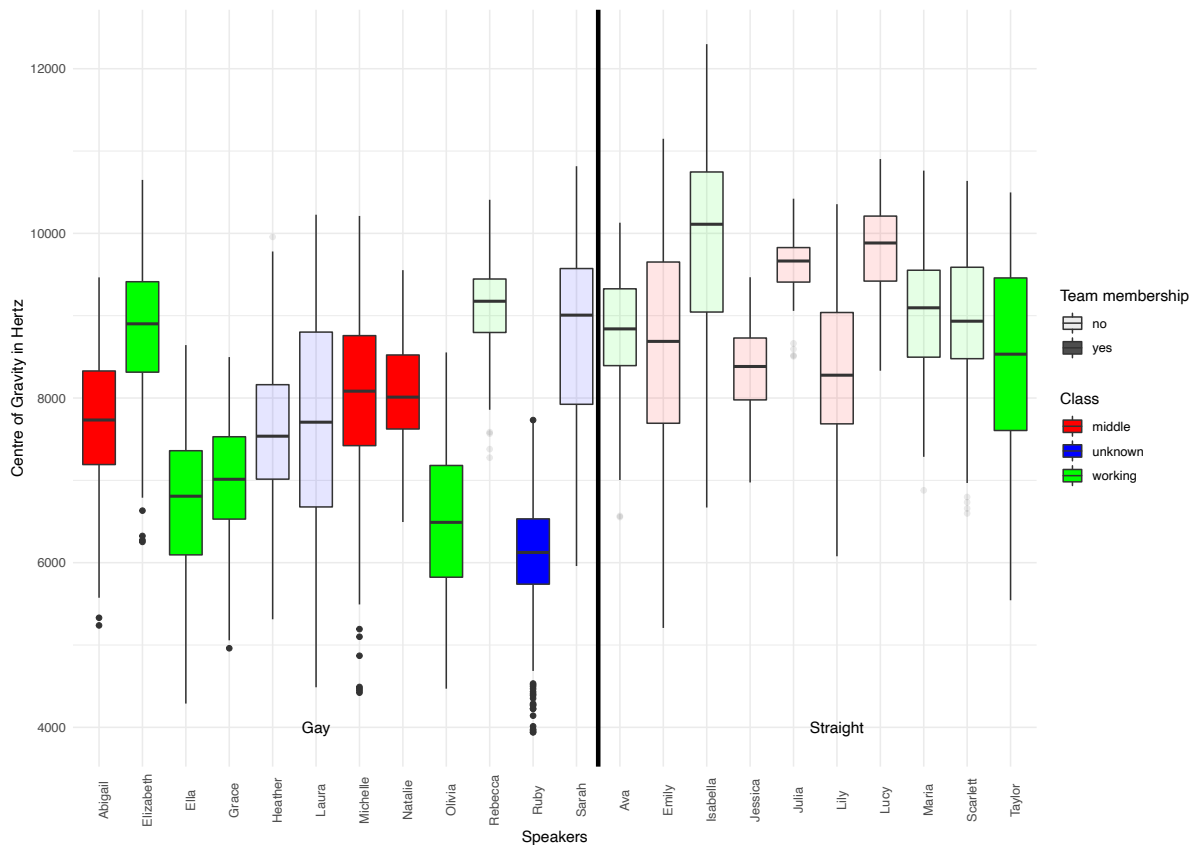
Figure 7 CoG observations (in Hz) for stressed syllable onset /s/ tokens, phrase medial

As with the pitch production, all the following are mixed-effects regression modelling created in R (2008) using the "lme4" package and the "lmerTest" package was used to summarise the results and obtain a *p*-value. These models are built-up in the presentation of the results. The models initially consider one of the social identity categories, as well as syllable placement and stress. Once these are presented, larger models are presented that consider multiple social factors and more nuanced details of the linguistic environment the /s/ token was in.

The variation according to sexual orientation is statistically significant, as confirmed by the mixed-effects regression model, the results of which can be seen in Table 26.

75

Ambisyllabic tokens were excluded due to their low occurrence and to allow for a binary comparison between coda and onset syllables.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	7388	232.93	31.72	0.000
Sexual orientation (straight)	1345.3	344.3	3.91	0.000872
Syllable (onset)	203.86	28.5	7.152	0.000
Stressed (unstressed)	-209.07	34.71	-6.024	0.000

Table 26 CoG regression model according to sexual orientation

Social class does not appear to affect /s/ realisations to a significant degree. Table 27 shows the results of the regression model according to social class. These results exclude tokens labelled as ambisyllabic and speakers whose social class was unknown.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	8275.06	361.14	22.91	0.000
Social class (working)	-201.45	483.72	-0.416	0.683
Syllable (onset)	200.84	29.46	6.818	0.000
Stressed (unstressed)	-194.66	35.99	-5.408	0.000

Table 27 CoG regression model according to class

As with sexual orientation, team association is also significant in CoG measures in /s/ realisations, along with syllable placement and stress. The results of the regression model according to team association can be seen below in Table 28.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	8589.54	213.41	40.248	0.000
Team association (yes)	-1442.37	332.06	-4.344	0.000315
Syllable (onset)	203.84	28.51	7.15	0.000
Stressed (unstressed)	-209.24	34.71	-6.029	0.000

Table 28 CoG regression model according to team association

As team association is shown to be more significant in influencing /s/ realisations, a regression model was fitted to consider the interaction between social class and team association, as well as syllable placement and stress Table 29 shows the results.

Fixed effects	Estimate	Std. error	t	p-value
(Intercept)	8751.67	311.91	28.058	0.000
Class (working)	148.99	440.05	0.339	0.74
Team association (yes)	-1270.69	507.92	-2.502	0.0254
Syllable (onset)	200.93	29.46	6.82	0.000
Stressed (unstressed)	-194.76	35.99	-5.411	0.000
Class (working): team (yes)	-383.2	671.86	-0.57	0.5775

Table 29 CoG regression model for /s/, according to social class and team membership

Along with the identity factors, another important influence on the CoG values was phonological environment. Table 30 presents the social and linguistic factors considered in the final model in this chapter. Team association was shown to be more influential than sexual orientation and therefore was used in this large model to consider with social class. The undetermined social class group was included in this model to account for all of the data, but again it is important to remember that they do not make up a cohesive group. Ambisyllabic syllables were also included to account for all the data.

Factors		
Social	**Linguistic**	
Class	**Stress**	
middle	stressed	unstressed
unknown	**Syllable**	
working	ambisyllabic	onset
Sexual orientation	coda	
gay	**Phrase**	
straight	final	medial
Team association	initial	
no	**Previous**	
yes	approximants	front vowel
	pause	fricative consonants
	back vowel	nasal consonants
	central vowel	plosive consonants
	diphthong	/r/
	Following	
	approximants	consonant clusters
	pause	fricative consonants
	back vowel	nasal consonants
	central vowel	plosive consonants
	diphthong	/r/
	front vowel	

Table 30 Social and linguistic factors for regression model

Table 31 clarifies which phonemes were included in the linguistic environment factors. These phonemes are coded using the ARPAbet used in FAVE forced aligning.

Linguistic levels for regression models					
Previous sounds			**Following sounds**		
Approximants			**Approximants**		
L			L	W	Y
Pause			**Pause**		
sp			sp		
Back vowel			**Back vowel**		
AA0	AH1	UH1R	AA0	AH1	AO2
AA0R	AO0	UW0	AA1	AH2	UH0
AA1	AO1	UW1	AA1R	AO0	UH1
AA1R	AO1R	UW1R	AA2	AO1	UW1
AA2	UH0	UW2	AH0	AO1R	UW2
AH0	UH1		**Central vowel**		
Central vowel			ER0		ER1
ER0		ER1	**Diphthong**		
Diphthong			AW1	AY2	OW1
AW1	EY1	OW2	AY0	EY1	OW2
AW1R	EY2	OY1	AY1	OW0	
AY1	OW0		**Front vowel**		
AY2	OW1		AE0	EH1	IH2
Front vowel			AE1	EH2	IY0
AE0	EH1R	IH1R	AE2	IH0	IY1
AE1	EH2	IH2	EH0	IH1	IY2
AE2	EH2R	IY0	**Consonant clusters**		
EH0	IH0	IY1	KR	KW	TR
EH1	IH1		**Fricative consonant**		
Fricative consonant			DH	HH	V
DH	TH	V	F	TH	
F			**Nasal consonant**		
Nasal consonant			M		N
M	N	NG	**Plosive consonant**		
Plosive consonant			B	G	P
B	G	P	D	K	T
D	K	T	**/r/**		
/r/			R		
R					

Table 31 Linguistic levels for regression model

Table 32 shows the regression model results for the two identity factors above, along with linguistic environments that have been detailed.

Fixed effect		Estimate	Std. Error	*t*	p-value
(Intercept)		8609.32	327.16	26.315	0.000
Class					
	unknown	-864.92	500.1	-1.73	0.1029
	working	149.96	433.09	0.346	0.73366
Team association (yes)					
		-1269.91	**499.87**	**-.2.54**	**0.021815**
Syllable					
	Coda	55.6	84.68	0.657	0.511531
	Onset	**218.49**	**84.51**	**2.591**	**0.00966**
Stressed (unstressed)					
		-224.2	**34.15**	**-6.566**	**0.000**
Phrase					
	Initial	64.84	38.43	1.687	0.091637
	Medial	39.08	25.22	1.55	0.121239
Previous					
	back v	-62.75	67.61	-0.928	0.353413
	central v	-27.6	78.92	-0.35	0.726593
	diphthong	**-148.14**	**73.86**	**-2.006**	**0.044968**
	fricative c	-57.6	88.03	-0.654	0.512898
	front v	-81.75	68.82	-1.188	0.234943
	nasal c	22.96	69.98	0.328	0.742875
	pause	-142.08	73.22	-1.94	0.05239
	plosive c	85.44	67.09	-1.274	0.202877
	/r/	**-319.68**	**161.64**	**-1.978**	**0.047982**
Following					
	back v	**106.55**	**49.31**	**2.161**	**0.030756**
	central v	**149.66**	**92.57**	**1.617**	**0.106162**
	c clusters	**-877.65**	**107.79**	**-8.142**	**0.000**
	diphthong	**271.22**	**62.39**	**4.347**	**0.000**
	fricative c	**339.73**	**56.47**	**6.016**	**0.000**
	front v	**425.65**	**52.24**	**8.148**	**0.000**
	nasal c	**238.2**	**66.76**	**3.568**	**0.000361**
	pause	**150.47**	**52.7**	**2.855**	**0.004308**
	plosive c	-16.01	46.21	-0.346	0.729099
	/r/	**-715.15**	**154.98**	**-4.615**	**0.000**
Class (unknown): team (yes)		-796.89	934.91	-0.852	0.406596
Class (working): team (yes)		-387.37	661.21	-0.586	0.566149

Table 32 CoG regression model according to all factors, significant factors emboldened

As can be seen in the results shown in Table 32, syllable, stress, and following sounds appear to have a significant influence on CoG measurements, as well as speaker's membership of the YTL.

3.4 Discussion

The following section will now consider the patterns that emerge from the data and argues the different identity factors may influence the investigated phonetic features.

3.4.1 Qualitative responses

As discussed in the introduction and the literature review, a gay voice for women is not as commonly stereotyped as the gay voice for men (for literature on the gay voice for men, see Gaudio, 1994; Smyth and Rogers, 2003; Levon, 2016). The lack of a stereotyped gay voice for women is confirmed by the comments provided by the participants. Many did not express the sense that a female gay voice exists and said it was not something they had considered previously. This lack of public awareness justifies further investigation into how women may index their sexual orientation, be it through dress, discourse, or phonetic features below the level of conscious awareness.

It was also seen within the group of gay members in the Yorkshire Town Ladies that they did not consider themselves activists within the LGBTQ community and expressed opinions that they did not need to "shout from the rooftops" about their sexual orientation (a quote from Natalie's interview). This presents an opportunity for further study with participants who commonly participate in the LGBTQ activist movement and frequently campaign for LGBTQ rights. While there are significant differences between the gay and straight participants in respect of their use of the phonetic features in this study, there may be still further differences between groups of gay female speakers who are actively part of an LGBTQ community which forefronts this particular identity. At the time of writing, there is no current sociophonetic research that takes into account a level of activism or inclusion in an LGBTQ community. There is research that focuses on communities of practice that are composed chiefly of gay women, such as Jones's (2011) study of a hiking group,

but such an approach has not yet been taken in the sociophonetic work. Further studies of communities of practice that emphasise sexual orientation would contribute in a significant way to the overall body of research on gay female speakers.

One of the interesting patterns to come out of the interviews was the movement away from the term "lesbian" and a shift towards "gay" being used for everyone that is attracted to members of the same gender. This preference for *gay* was also noted by Sauntson and Morrish (2012, p. 167), in which the researchers write, "[we] may speculate that to identify as lesbian would seem to be too determining to these young women". As the participants in both studies are predominantly under 35 years old and all are British, it may be that there is a shift away from the word "lesbian" for younger British women.

However, as noted in the results, there is still a reliance on "lesbian" in more formal situations, for instance when filling in forms. To place this in a wider UK context, there was public outcry when the word "lesbian" was not used to describe a famous historical figure, Anne Lister, on a recently unveiled plaque in York (BBC 2018). While Lister was described as "gender non-conforming" and not "gay", there was a call to specifically include the word "lesbian" on the plaque, as those protesting against the wording wanted it clearly stated that Lister was a woman who was attracted to other women. While these three small-scale examples may not be enough to definitively show that there is a shift to new terminology preferences by younger gay women, they do demonstrate the need for further research.

3.4.2 Mean Fundamental Frequency

The first noticeable factor about the mean F0 data for this group of speakers is the fact that the speakers use a lower mean F0 than that reported in much of the previous work cited. In Van Borsel *et al.*'s (2013) study of Dutch speakers, the gay participants had a mean F0 of 194.5Hz and the straight speakers a mean F0 of 204.4Hz. In the study by Rendall *et al.* (2008) of Canadian English speakers, the gay speakers a mean F0 of 198Hz and the straight speakers had a mean F0 of 202Hz. The mean F0 values for the present study are also much lower than the

averages presented by Simpson (2009) and Traunmüller & Eriksson (1995), which both cite female speakers as having a mean F0 of 200-220Hz. The gay speakers in the current research have a mean F0 of 181Hz and the straight speakers a mean F0 of 186.5Hz.

Sexual orientation was found to be a significant factor in respect of mean F0 across the present group of speakers. The difference between gay speakers and straight speakers is not large, with the regression model estimating only a 5.5Hz difference between the groups. However, even after accounting for individual speaker differences, sexual orientation appears to be a significant influence on mean F0 production. This matches Van Borsel *et al.*'s (2013) study, which also found a significant difference in F0 between the gay and straight Dutch speakers.

However, this data also shows that there is more diversity within the group of gay speakers than differences between the sexual orientation groups, and even more variation than within the group of straight speakers. The straight speaker with the lowest mean F0, Isabella, averages 174.4Hz, and the highest mean F0 value for the straight speaker, Lucy, is only 28.6Hz higher, with a mean of 203Hz. The gay speaker with the lowest mean, Natalie, averages 154.5Hz and the gay speaker with the highest mean, Grace, has an average F0 rating of 209.4Hz. This is a difference of 54.9Hz, which is 26.3Hz larger than the range for the straight speakers. While gay speakers tend to have a lower pitch than the straight speakers, on the whole, it is clear that there is distinct diversity within the group.

This within-group difference is similar to that reported by Rendall *et al.* (2008), who did not find a significant difference between their groups of gay and straight female speakers, but cited considerable variation within the groups themselves. They also found that the speakers with the five lowest mean F0 measurements were all gay, and the speaker with the highest mean F0 was also gay. Rendall *et al.* and the present study continue to demonstrate the significance of in-group diversity and the importance of researchers considering other social factors when comparing groups.

Social class is also a significant factor in the data for this set of participants. Middle-class speakers have lower mean F0 measurements than the working-class

group. These differences are statistically significant across the class division. As with sexual orientation, the differences between the groups are not great.

Being a member of the Yorkshire Town Ladies team also appears to have a significant effect on F0 values. Members of the team generally have lower F0 measurements than those speakers who are not part of the team. While the team was beginning to fray by the start of research, these results hint at the fact that it may justifiably have been regarded as a community of practice (Eckert and McConnell-Ginet, 1992) at one point, and there may still be signs of the influence of association with the team. This CoP may have used certain phonetic features to index identity traits that will be considered further in Section 3.5.

In the mixed-effects regression model that accounts for social class and team association there does not appear to be an interaction between these two factors, as neither was significant. However, the importance of being on the team will be discussed further in 3.5.

3.4.3 Fundamental Frequency Range

Fundamental frequency range is one of the few features that has been studied with regard to female sexual orientation, and therefore was important in the current study. Moonwomon-Baird (1997), Waksler (2001), and Van Borsel *et al.* (2013) consider pitch range in their research on gay and straight women. While Moonwomon-Baird cited pitch range as an important factor in distinguishing her groups of speakers, Waksler found that the difference between the gay and straight speakers in her study was not significant, and that there was extensive variation within the groups. By contrast, Van Borsel *et al.* (2013) did find that there was a significant difference between the F0 ranges used by their gay and straight participants.

The current research supports Waksler's findings in respect of F0 range and sexual orientation. While the gay speakers exhibit a slightly lower mean range in semitones, this difference is not significant when considering the F0 observations across the entire recording. There is also variation within the groups of speakers. Ella has a F0 range of 5.2 semitones, while Heather has a F0 range of 13 semitones;

both identify as gay. Similarly, the straight speaker with the narrowest range is Lucy, who has an F0 range of 7.7 semitones, while the straight speaker with the widest range is Ava, who has a range of 14.4 semitones.

It appears that pitch range is not an important factor when considering sexual orientation more broadly. This is also confirmed when only considering the retelling portion of the recording. As mentioned in the results presented above, it is possible that parts of the interviews for some speakers could have led to more enthusiastic conversation than it did for others. This enthusiasm may have resulted in a "livelier" intonation, which could increase the speaker's F0 range. For this reason, it is important to consider only the retelling portion of the recordings, as they are more controlled. Recording all of the speakers discussing the same topic would promote comparability of their F0 ranges. This was also a feature of Waksler's (2001) work, in which she asked all the participants to retell the story of the *Wizard of Oz*.

The pitch ranges for all the speakers, regardless of group, are slightly lower in the retelling task. However, even when only considering the retelling portion, there is still no significant difference between the gay and straight speakers. This feature is, therefore, worth considering further, as there have been two studies in which F0 range was found to be significantly different between gay and straight speaker groups (Moonwomon-Baird, 1997; Van Borsel, Vandaele and Corthals, 2013) and two studies in which F0 range was found not to be significantly different (Waksler, 2001 and the present study). The difference in range could be due to other social factors that have not been thoroughly analysed to date, but will be considered further in Section 3.5.

Social class also is not significant in terms of F0 range differences. Both in the entirety of the recording, and also the narrow view of just the retelling, social class does not have a significant effect in this group of speakers.

The only difference in pitch range appeared in relation to team association. When considering the entirety of the data, participants who were not on the team had an average pitch range of 10.2 semitones, while those on the team had an average of

7.8 semitones; this difference was found to be significant. However, when only considering the retelling portion, it is no longer significant. The team average for the retelling portion drops slightly to 7.3 semitones, while the range for non-team participants drops to 9 semitones. There is also no significant difference in retelling pitch range according to team association.

As with the general F0 measurements discussed above, there are preliminary signs that there may be an influence of team association, but these findings do not hold true across all contexts. While the YTL was a team that had been near the end of their regular interactions when the recordings took place, I believe the evidence suggests that at one point there may have been stronger community ties that could have led to more clearly demarcated in-group speech, in particular around key identity features.

3.4.4 /s/

The second phonetic feature to consider is the voiceless alveolar fricative. As discussed in the literature review in Section 2.5.2, fricatives in general have been demonstrated to be an important set of sociophonetic variables in indexing gender and sexual orientation. In particular, /s/ has been prominent in the previous research and in the present study it appears again to play a significant role.

Based on an initial inspection of the data, the centre of gravity (CoG) for gay speakers generally appears to be lower than that for the straight speakers. This difference is statistically significant when sexual orientation is considered alone, a finding which matches the results of Podesva and Van Hofwegen (2016), who found that gay female speakers had lower CoG measurements than either their straight country women and their straight town women. Similar results were also found by Hazenberg (2012) when comparing queer women with straight women. Based on these three studies of English speakers in three different countries (England, Canada, and the USA), it appears that CoG may be a significant factor when considering a gay voice for female speakers.

When considering the separate regression models for the participants, as well as the larger model presented in Table 32, it is clear that team association is a significant factor in influencing /s/ realisations.

Along with the social factors discussed above, internal linguistic factors are also significant in /s/ production, as would be expected. Syllable placement and stress are consistently significant factors across all /s/ observations, a trend which was also found in the studies by Hazenberg (2012), Holmes-Elliott & Levon (2017), Podesva and Van Hofwegen (2016), and Zimman (Zimman, 2012). This data shows that the following phoneme appears to have a stronger influence on the /s/ CoG measurements than the preceding sound.

In Table 32, almost every following sound has a statistically significant effect; among the preceding consonants only preceding /r/ has a significant effect. The occurrence of /s/ in a consonant cluster (such as /str/, /skw/, and /skr/) has one of the largest effects in CoG measurements, and estimates a CoG value that is 877Hz lower.

The observations made on the basis of the /s/ measurements show that /s/ productions are being used in identity work and that there is variation based on group association. While the linguistic factors in this are important and have an impact on CoG, group connections are also significant in this data.

3.5 Sporty participants

Based on the results from this study, I propose that some participants are signalling a sporty identity through phonetic features. While sport and football may not be the particular catalyst for this persona, without further ethnographic work it is not possible to make any assumptions on other factors that the participants may be signalling. For the present study, sport will be considered the nucleus for the specific identity presented.

It is possible that some of the participants identify with a more "butch" gender identity, which is commonly associated with gay women. Levitt and Horne (2002)

write that the dichotomy of butch-femme identities began in the USA as early as the 1950s. Butchness is often associated with masculinity, but as Jones (2018, p. 2) writes, "A butch lesbian, therefore, is not necessarily role-playing a male identity, but projecting a *lesbian* one; this is a way of performing a version of womanhood that indexes difference from normative expectations of femininity" (emphasis in original text). While ideas of butch and femme are frequently considered when working with gay women, not everyone identifies within this dichotomy (Levitt and Horne, 2002).

The gay participants in the present study were asked about the concept of butch and femme and none said she particularly identified with either. A few, including Elizabeth, Michelle, and Ruby, said they felt that if butch and femme identities were the poles of a scale, they would fall somewhere in the middle, with some days leaning more or less to one side, either based on mood or the social expectations of a specific context. However, not every YTL member gave this description. Therefore, I argue that there may be more of a shared identity around participating in sport, as opposed to a specific butch identity.

As Sauntson and Morrish (2012, p. 153) argue in their study of a university women's football team, by playing sport the team members are inherently challenging gender norms. They write, "[p]articipants in women's football are likely to be women who are less concerned than some others about observing the norms of hegemonic femininity in this age group, including heteronormativity." Caudwell (2007, p. 184) also writes, "Analyses demonstrate that football in England is gendered as masculine and racialised as white". This view of football as a male sport was also shared by some of the YTL members. Olivia and Taylor both discussed the association of certain sports with specific genders; football was deemed a man's sport, stereotypically. This correlation between football and masculinity, therefore, means that any woman playing football is necessarily challenging gender norms to some extent.

This perception of football being a masculine sport had an impact on how much YTL members could access playing, particularly as teenagers. Many of the YTL discussed the lengths they had to go to in order to play football as young adults

because there were no teams for girls after a certain age. Michelle explained that boys and girls were allowed to play football together until the age of 11; after that point they had to play on separate teams. However, as there was not as much interest in girls' football, either among the children and those in charge of resources, girls above the age of 11 frequently did not have a team to play on. Elizabeth also experienced the loss of football around 11, when her parents decided it was no longer acceptable for her to play and instead encouraged her to take up other sports which were deemed as more appropriate for her gender. Michelle and Ruby both discussed that girls were not allowed to play football in physical education classes in secondary school. They both found this unacceptable and continually asked to play with the boys until they were allowed to practise on their own, in the case of Michelle, or play with the boys, in the case of Ruby.

Ella, Michelle, and Ruby also comment on a desire to be seen as equally talented as any of the boys playing football, if not more talented. Each one experienced being the only girl playing on a team of boys, either casually or competitively, and they expressed a desire to "prove themselves". This signifies that football and sport were an important part of their identity and something that many team members put effort in to continue to play after they were told to stop on grounds of their gender.

This potential disregard for gender norms and distancing from hegemonic femininity may be partially signalled through voice qualities, which led to the distributions of mean F0, F0 range, and /s/ productions seen in the YTL members. While this sporty identity may have some similarities to a butch identity, as both are marked by a shift away from typical gender norms, the participants in this study do not appear to have a conscious, targeted desire to be seen as butch. Instead they have invested time in their youth to participate in football, for some even if that has meant being the only girl on the pitch.

Taylor, who is the only participant that is part of the YTL and straight, is worth considering further. She tends to pattern with other straight participants and non-team members. She has a higher mean F0 and a more fronted /s/ than many of the gay YTL participants. Taylor also acknowledged playing football from a young age and throughout university. She joked about being one of the few straight women on

her university football team. However, despite this connection to sport, she does not seem to share the sporty phonetic features of the other team members. This indicates that there is a specific connection not only with being sporty, but also gay. While playing sport, and in particular a sport typically associated with men, may challenge hegemonic gender norms to begin with, this is compounded when that person also identifies as gay and may shift them even further away from typical gender presentation.

The intersection of sporty and gay is further seen when considering the gay participants that were not on the team. Sexual orientation was not significant in /s/ production when considering all the factors, but team association was. Similarly, the gay participants with the three largest pitch ranges for both the overall recording and the retelling portion in particular, are all not members of the YTL. As Taylor, the straight member of YTL, does not always pattern with the rest of the team, and the other gay participants do not always match the production of the gay members of the YTL, this suggests that there is a more specific identity for gay members of the YTL.

3.6 Conclusion

This study has provided important insights into the speech production of this group of speakers and begins to hint at variation that may occur beyond the group. Social class has a minimal effect on variation in the present study for F0 range and /s/ production. The only case in which social class was significant was for mean F0. It is possible that speakers may be indexing social class in other features of their speech, which were not investigated in the current study.

F0 ranges appear to be of little consequence in performing a gay identity among these white British English speakers. However, it is still worth investigating the speech of other demographic groups and with other groups of speakers based on the previous work that found systemic variation in F0 range.

However, mean F0 does appear to be a phonetic feature that speakers may use to do sexual identity work. Similar findings are reported by Van Borsel *et al.* (2013) and

to a lesser extent, Rendall *et al.* (2008). Therefore, there is a need to continue researching this area to investigate how variation in mean F0 as a correlate of sexual orientation interacts with variation connected to geographic, ethnic, and group identity factors.

The production of /s/ also appears to be significant when considered alongside sexual orientation among the Yorkshire Town Ladies. This matches previous research by Hazenberg (2012), Podesva and Van Hofwegen (2016), and Zimman (Zimman, 2012), all of which found /s/ to be significant in their studies of gender and sexual orientation. Further research into the extent of this variation would help us to understand how /s/ may be used to index sexual orientation, even as a feature that is not above the level of consciousness for most listeners.

From the qualitative data and the quantitative data, it is clear that phonetic variation is not a correlate of sexual orientation alone, but also of other identity factors that are convolved. Team membership was one of the most important identity factors in the quantitative data, and as can be seen through the qualitative results presented, there were many views of sexual orientation that the team shared that other gay participants who were not on the team did not share. This research shows that there is no determinist link between sexual orientation and phonetic features, but the variation happens between certain groups of gay and straight speakers. This thesis has presented the argument for a sporty gay identity that accounts for the variation within the group of gay speakers.

With an understanding of how speakers produce these phonetic features, it is important now to investigate how listeners perceive some of these features. The following chapter will present the perception study focusing on sexual orientation.

Chapter 4: Perception

4.1 Introduction

The second element of this PhD was to consider listener perceptions of a gay voice for women and to investigate if there were any consistent patterns underlying what listeners might perceive as "sounding gay" for a female speaker. It was particularly important to connect this to the production study described previously, as it would allow for a comparison between what listeners may expect and what speakers actually do.

In order to obtain this comparable data, it was decided that speaker pitch would be an ideal place to begin the investigation of a perceived female gay voice. Pitch was one of the few phonetic features the participants in the production study cited as being potentially indicative of a gay voice. Pitch has also been studied previously in association with language and sexual orientation (Moonwomon-Baird, 1997; Waksler, 2001; Rendall, Vasey and McKenzie, 2008; Van Borsel, Vandaele and Corthals, 2013; Sulpizio *et al.*, 2019) and appears to be a significant feature when discussing a female gay voice.

The research questions for this study are as follows:
1. Do participants perceive a voice with lower F0 to sound more homosexual than a voice with a higher F0?
2. Do perceptions of femininity and homosexuality negatively correlate?
3. Do different groups of listeners perceive homosexuality differently?

The survey was presented to listeners in the guise of a research project on perceptions of Yorkshire accent, rather than specifically acknowledging the researcher's interests in gender and sexual orientation. If participants were too aware of the focus on sexual orientation, it would be possible that they might overthink their responses and potentially provide responses that were based on

societal expectations, and not based on their individual beliefs. Levon (2006) also withheld the true focus of his research from his participants for a similar reason. It was hoped that by focusing the study on a regional accent that many British participants would be familiar with, questions on sexual orientation embedded within the experiment would be less conspicuous. As participants would be relying on their knowledge of Yorkshire accents (and maybe Yorkshire people), they might not pay such close attention to an unexpected quality they are asked about.

Regional accents are frequently discussed in British culture (Beal, 2010, p. 1). They are regularly commented on in the media and laypeople typically have a plethora of opinions on regional accents. As Yorkshire is the biggest county in England, its accent/dialect is also one of the most commonly discussed regional varieties. This awareness of regional varieties is much stronger than an awareness of a gay voice for women, to judge by the amount of linguistic literature alone. By presenting the study as a focus on regional accents, people might be more inclined to participate and might have strong, consciously-held opinions before even starting the survey. This would likely not be the case for gay voices for women.

Additionally, by framing the survey as a view of Yorkshire, it would then be possible to use a speaker from Yorkshire to obtain the stimuli for the survey. As the production study had already sought female participants from Yorkshire, there was already a protocol for finding female speakers from the area. The selection of the individual speaker to provide the sentence stimuli will be discussed in more detail in the methodology below.

The final benefit of using a speaker with a Yorkshire accent was the associations of "roughness" or "hardness" that listeners have with Northern UK accents (Wales, 2000). As Isabella noted in her interview, described in Section 3.3.1.3, she believed that some might associate her voice with a female gay voice because it is perceived as "rough" simply by being a Yorkshire accent. As has been cited by the participants in the production study, female gayness is often associated with masculinity and these "rough" accents might then be connected to a female gay voice.

It is of course not being argued here that gay women are intrinsically masculine, or that northern English accents are associated with a gay identity for women. However, it is possible that these two separate stereotypes may interact and that listeners could be combining these associations. The focus on the current study is how participants perceive the quality of "homosexuality" and what a gay voice might sound like.

Before presenting the study, it is important to note how terminology is henceforth to be used. As has previously been discussed, this thesis has used the word "gay" when describing people who are exclusively interested in romantic and sexual relationships with people of the same gender. This has been done based on participant preferences, discussed in Section 3.3.1.1. However, in this section the word "homosexual" is used frequently when describing the quality participants were asked to rate. The reason for this change in terminology is one of clarity, and is discussed in detail below in Section 4.2.2. The thesis uses the word "gay" when discussing *people*, but uses the word "homosexual" when discussing the *trait* participants were asked to rate.

4.2 Methodology

While other perception studies have created the listening stimuli from the speech of a mixture of self-identified gay and straight speakers (Munson, Jefferson and McDonald, 2006; Maegaard and Pharao, 2016; Sulpizio *et al.*, 2019), this study created the stimuli from the utterances collected from just a single speaker. The previous studies have found differing levels of significance in participants' perceptions of sexual orientation in others, or "gaydar", but have then analysed multiple segmental and suprasegmental features in order to investigate what participants may have been responding to in their ratings. For example, Sulpizio *et al.* (2019) analysed 27 different segmental and suprasegmental measures for all their speakers, based on three different languages.

Instead, the present study follows a similar methodology to that used by Levon (2006), using utterances from one speaker that have been digitally altered as stimuli for listeners to rate. This method allows for a more targeted investigation that can

study how participants respond to a single feature, in this case perceived pitch. While the previous work has provided valuable insight into how accurately listeners perceive sexual orientation, this study was more concerned with how sensitive listeners are to a key feature and whether this feature impacts perceptions of sexual orientation.

Also similar to Levon (2006), as well as the studies listed above, this perception study was designed to explicitly ask participants about their attitudes to the qualities presented. As will be discussed below, multiple qualities were included in the experiment in an attempt to draw attention away from sexual orientation specifically. However, participants were still asked explicitly to respond to each quality. This was done to allow for more direct comparisons to previous work and to build on the foundation that has been laid for this type of research.

These direct questions about various qualities could create a bias in the data, as people may not be comfortable rating qualities based on a short sound clip of a stranger. This potential discomfort is discussed in detail in Section 4.4.1. However, this study presents an interesting insight into the attitudes people are willing to admit explicitly, even for topics they may find uncomfortable. This is a good starting point for research that is still relatively new and uncommon when considering female sexual orientation. By having this initial understanding of participants explicit views, a more informed implicit experiment can be designed in the future.

4.2.1 Stimuli collection

The sentences read by the speaker to create the survey stimuli come from the Intonational Variation in English (IViE) Corpus (Grabe, Post and Nolan, 2001). This corpus material was specifically chosen because it has a high proportion of voiced sounds. As the IViE project focused on speaker intonation, voicing was a priority for their study. Similarly, this current study is also focused on voiced sounds, as F0 values were to be manipulated, and for this reason these sample sentences would be ideal. There are 22 sentences from the corpus that include both statements and questions. The sentences from the IViE project were put in a random

order and presented on a laptop screen for the speaker to read. For a list of the 22 sentences, see Appendix 6.

The speaker was selected for the current study from the pool of participants in the previously described production study, and there is referred to as Scarlett. She had already participated in the production study interview before she was approached to produce the stimuli for the perception study. This particular participant was selected because she had an average F0 near 200Hz. As 200Hz is frequently cited as the lower-frequency end of adult female speakers' range (Traunmüller and Eriksson, 1995; Simpson, 2009), she was an ideal candidate. This would allow for the opportunity to lower and raise her F0 artificially while keeping the stimuli sounding natural. If a speaker was chosen that had a mean F0 of 160Hz, for example, there would not be as much of an opportunity to decrease her mean pitch and maintain a natural-sounding female voice.

Scarlett was also chosen because she has a more pronounced Yorkshire accent than some of the other participants in the production study, and as the survey was being presented as a study of Yorkshire accents, this was important. The features of a Yorkshire accent were described previously in Section 2.4. Scarlett having an identifiable Yorkshire accent means that she frequently uses common features associated with it, which include no distinction between the FOOT and STRUT vowels, /a:/ realisations for the BATH vowel, and monophthongal realisations for the MOUTH, GOAT, FACE, AND PRICE vowels (Beal, 2010). While all the participants in the production study were from Yorkshire, the participant chosen to produce the stimuli for the perception study used these Yorkshire features more frequently than some of the others, and was deemed the best choice for the perception study based on accent and mean pitch.

Scarlett identified as straight and she was aware of the aim of the study, but not of how her voice would be manipulated. She was recorded on a Zoom H4n solid-state recorder (sampling rate 44.1 kHz; bit depth 16-bit) while wearing a Shure SM-10A (dynamic cardioid) headset microphone.

Since there was no way to control the playback quality participants experienced when completing in the survey, it was important that the quality of the sound files was as clear as possible before distribution. To increase the volume and decrease ambient noise, the original sound file was filtered using Praat (Boersma and Weenink, 2019) and Audacity (2019). First, in Praat, the recording had the peak scaled to 0.8 in order to create a new absolute peak, which would increase the amplitude of the entire file. However, this increased the amplitude of the acoustic energy in the file, including the background background noise. To remove the white noise, the file was then filtered in Audacity using the Noise Reduction function. A portion of the background noise from a non-speech-active portion of the recording was selected and the entire file was filtered by reducing it by 18 dB, at a sensitivity of 6, and a frequency smoothing (bands) setting of 3. This reduced the white noise, which vastly improved the listening quality for the entire sound file, and it was this edited sound file that was then segmented and used in the survey.

The recording was segmented into 22 individual sound files containing one sentence apiece, which could then be further manipulated in Praat. Each file was run through a Praat script (Fecher, 2015) that first assessed the mean F0 for the duration of the file, and then altered it. This script can be viewed in Appendix 7. Each sentence was scaled down by 50 Hz, 40 Hz, 30 Hz, and 20 Hz and was also scaled up by 50 Hz, 40 Hz, 30Hz, and 20 Hz. This produced a total of 176 altered sentences from the originally read 22 sentences. Pooling the altered and unaltered sentences brought the total number of candidate stimuli to 198.

While there is a larger perceptual difference in decreasing the stimuli by 50Hz than increasing it by 50Hz, participants still reported a sensitivity to an F0 increase in much the same way as an F0 decrease. This will be seen more clearly in the results.

4.2.2 Survey creation

Having collected the 198 sentences, it was important to select 40 that would be included in the final survey. Each manipulated sentence was first evaluated for naturalness. There was a particular focus on sound files that were increased or decreased by 40Hz and 50Hz, as they experienced the largest transformation and

would possibly have the least natural sound. Once sentences were rated for naturalness (either "natural" or "unnatural") by the researcher, a collection of the files was selected for a testing on a small group of listeners.

In order to ensure the manipulated sentences could pass as natural to listeners, a pilot study was devised that focused on the perceptions of "naturalness" among a small group of listeners. The pilot survey was created using the online platform Qualtrics (2018; Provo, UT) and included 20 sentences to be rated on a five-point Likert Scale of "Not natural at all" to "Completely natural". As I was particularly focused on sentences that were increased or decreased by 40Hz and 50Hz, they were favoured in this initial survey. Sentences that had not been altered as drastically were deemed to be more natural to listeners and therefore did not need to be thoroughly piloted as did the sentences increased or decreased by 40 or 50Hz. A small selection of stimuli sentences from the larger pool that were deemed natural-sounding by the researcher and her supervisors were not included in the pilot study.

The pilot survey was distributed online, and participants were told that they would be listening to a series of sentences that contained a mix of natural and digitally altered files. They were asked to decide how natural the voice sounded. A breakdown of the manipulated sentences that participants listened to can be seen in Table 33.

F0 Manipulation	Number of sentences
Decreased 20Hz	1
Decreased 30Hz	2
Decreased 40Hz	2
Decreased 50Hz	3
Natural	4
Increased 30Hz	2
Increased 40Hz	4
Increased 50Hz	2

Table 33 Manipulated sentences in pilot survey

There was a total of 11 participants that took part in this pilot survey to judge the sentences that would be included in the final research. Participants were friends and family of the researcher and none of them had previous linguistic training. Based on their naturalness ratings, some of the sentences were included in the larger survey and others were not included.

Due to the short length of each sentence, and the survey directions encouraging participants to answer as quickly as possible, it was decided that 40 sentences would be an acceptable number to present to participants in the full study. This would allow for a large sampling of sentences from all of the pitch manipulation tiers. As with the pilot survey, this large-scale survey was created online using the website Qualtrics. This would allow participants to respond to the survey either online or on their mobile devices. By making the entire survey digital, participants were able to respond at any time and anywhere they could access the Internet. This allowed for a broad range of responses from participants who were unable to physically access the University of York.

Although only one speaker's voice was represented in the stimuli, participants were led to believe that they were rating multiple speakers. For this reason, it was important to ensure that there were not too many examples of the same read sentence from the IViE corpus, and the ordering of the stimuli would also be crucial. Each sentence from the IViE corpus was only used a maximum of two times so listeners would not be able to hear the similarity in the reading, despite the pitch alterations.

Though only 20 sentences were tested in the pilot study, these ratings guided the selection of the 40 stimuli sentences that were used in the full survey. While the pilot study did not include many alterations of 20 or 30Hz, more of these stimuli were included in the full survey. They were deemed more natural-sounding since they did not deviate as far from the original recording. Other stimuli were included in the larger survey that did not appear in the pilot study, as they were deemed natural-sounding by the researcher and her supervisors without needing to be piloted.

Along with ensuring that there were not too many repetitions of the same read sentence, an effort was also made to have an equal number of each type of manipulation. Therefore, each manipulation tier was used four to five times. Table 34 outlines the number of stimuli sentences for each manipulation tier used in the survey.

F0 Manipulation	Number of sentences
Decreased 20Hz	4
Decreased 30Hz	5
Decreased 40Hz	5
Decreased 50Hz	4
Natural	4
Increased 20Hz	4
Increased 30 Hz	5
Increased 40Hz	4
Increased 50Hz	5

Table 34 Manipulated sentences in survey

As described above, it was important to ensure that the order of the presented stimuli did not allow participants to hear the same level of manipulation too many times in a row; for example there should not be a sentence decreased by 50Hz followed immediately by a second sentence that was also decreased by 50Hz. For this reason, the stimuli were presented in randomised chunks. Each chunk contained four or five sentences that had all been manipulated at different levels, and then these individual chunks were presented randomly to every participant. This method of randomised chunks ensured that there would not be any unintentional duplication, while still allowing for participants to hear a semi-random order of the data.

Importantly, the first sentence that every participant heard was the same. This sentence was one of the examples of the natural voice of the speaker. This was chosen to be first for every participant in the expectation that participants would subconsciously calibrate to that pitch. It is likely that all the participants would compare one sentence to the next to determine if one voice was higher pitched than another. By presenting the first sentence they heard as the natural voice of the speaker, the participants were starting from the same place. While it is unclear if participants would be able to "calibrate" to this voice, by all experiencing the natural sentence first, they were all given the same starting pitch to compare against.

With all the stimuli created and ordered in the way described, it was important to consider the traits the participant would be asked to give ratings for. Many of the qualities were chosen from previous perception studies (Hiraga, 2005; Levon, 2006; Campbell-Kibler, 2007; Klofstad, Anderson and Peters, 2012; McAleer, Todorov

and Belin, 2014; Mileva *et al.*, 2019), and these were used to conceal the questions about gender and sexual orientation. Qualities such as *friendly*, *intelligent*, and *trustworthy* were used. Participants were also asked to rate *feminine*, *homosexual (i.e. lesbian)*, and *low pitch* as the traits associated with a gay voice. Finally, participants were asked to rate how old they thought the speaker was to see if there was any correlation with the target qualities.

These questions were presented as seven-point Likert scales ranging from "strongly agree" to "strongly disagree". For an example of the final survey question used for each stimulus sentence, refer to Figure 8. For each of the stimulus sentences that participants were asked to respond to, there was a "force response" option selected, which means that participants were forced to rate each quality before they could move on to the next sentence. This was done to ensure that there would not be too many questions skipped that would then affect the statistical analysis later.

Rate the extent to which you think the speaker of this sound clip displays the following qualities
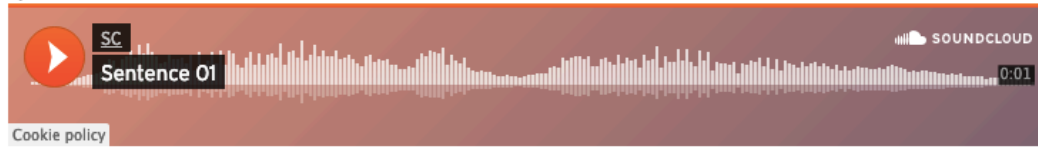
	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Friendly	O	O	O	O	O	O	O
Intelligent	O	O	O	O	O	O	O
Feminine	O	O	O	O	O	O	O
Trustworthy	O	O	O	O	O	O	O
Homosexual (i.e., lesbian)	O	O	O	O	O	O	O
Low pitch voice	O	O	O	O	O	O	O

Rate the age of the speaker

	Under 18	18-24	25-34	35-44	45-54	55-64	65 and older
Age	O	O	O	O	O	O	O

Figure 8 Survey questions as presented to participants

Careful consideration was given to how the qualities would be presented to the participants when asked to rate the sentences. Sulpizio *et al.* (2019) asked participants to rate stimuli on a Likert scale from "from 1 (*completely heterosexual*) to 6 (*completely homosexual*)" (p 7). However, for the present study it was decided this scale could lead to ambiguity since, as there are many other sexualities that exist, it is not clear if participants would accept heterosexuality and homosexuality as a dichotomy. Similarly, for some participants the inverse of "feminine" might be "masculine", but for others it might not be. In order to avoid confusion, with some qualities having dichotomies ("friendly" vs "unfriendly") and others not ("more feminine" vs "less feminine"), it was decided that the best practice would be to only rate the scale of the quality mentioned and not connect it with another quality.

102

A second consideration was how to present the question of sexual orientation. Research by Lick and Johnson (2016) demonstrate that there is a clear bias for participants to assume the majority of people are straight when asked to indicate a person's sexual orientation. In their research, even when told that 50% of individuals in question were gay, participants still overwhelmingly used the label of "straight" throughout the survey when assuming a person's sexual orientation. Owing to this "straight categorization bias", as Lick and Johnson call it, there was a concern in the present research that if the quality was presented as "straight" or "heterosexual", participants might rate this quality highly as a default bias. As a way to undermine this as much as possible, it was decided the quality participants would be presented with would be how gay someone is as a way to forefront this particular quality.

While the participants in the production study overwhelmingly agreed that they preferred the term "gay" instead of "lesbian" (see Section 3.3.1.1) and none of them used the term "homosexual" to describe themselves, there was a concern that participants in this survey may not be familiar with the use of "gay" to describe women. As this survey was designed to be sent to as many people as possible, across as many ages as possible, there was potential that not every participant would identify "gay" with female sexuality. For this reason, it was decided that *homosexual (i.e. lesbian)* would be the clearest and most accurate way to describe the quality that was being investigated because every participant would be familiar with the term "homosexual". In order to ensure that participants believed that all the sound clips were coming from a female speaker, "(i.e. lesbian)" was included.

The first information a participant would see when clicking the link to the study was a description of the study that states, "The purpose of this research is to add to the body of knowledge of women's speech and language." It was designed so that participants would always assume they were listening to a female speaker specifically. While they were not explicitly told how many different speakers they would be hearing, it was hoped that by not including that information, participants would assume they were either all different speakers or at least multiple speakers within the stimuli.

Before participating in the full survey that is described above, participants were given a chance to practise by filling in an identical question page but listening to a completely new voice. This allowed participants the chance to see what they would be asked to complete in the survey and to adjust their sound equipment accordingly before starting the full survey.

Along with the survey stimuli questions, the survey also included questions to gather biosocial data from each participant. This data was used to understand the identities of the participants and to allow for statistical comparisons between groups. Participants were asked their age, gender identity, sexual orientation, country of origin, region (if they were from the UK), whether they had any linguistic training, and if they had linguistic training, the extent of it.

For all of the questions apart from those relating to gender identity and sexual orientation, participants were asked to select an answer from a set of options. However, for gender identity and sexual orientation, participants were provided with an open textbox that they could fill in with their preferred answer. This openness allowed participants to identify themselves as accurately as possible, and also did not presuppose any identity or sexuality. However, some of these categories were collapsed into larger groups at the analysis stage, which will be discussed further in the results section.

Before distributing the survey, it was decided that a minimum of 60 participants would be adequate to run statistical analysis of the responses.

4.2.3 Survey distribution

While anyone was able to take the survey online, British participants were most desired as they would be most familiar with the Yorkshire accent and might be more perceptive of subtle variation in the voices. For this same reason, it was requested that participants be native speakers of English to maximise the probability that they would be sensitive to subtle phonetic differences in the test samples.

Friends and colleagues were asked to share the survey over their social media accounts. In order to target British speakers, people asked to share the survey were either British themselves or had strong ties to Britain, either through education or occupation. As discussed at the beginning of this methodology, it was decided that gender and sexual orientation should not be especially highlighted in this survey. For this reason, I did not personally distribute the survey through social media, as it might be connected to my online profile by would-be participants and they would be able to see my focus on gender and sexual orientation. While my name is on the project for participants to see, there was an attempt to distance my online research profile from the project itself.

Fliers were also distributed throughout York with information and links to the survey. These were posted around the university or given to friends and acquaintances either to take themselves or to distribute further. If I directly gave acquaintances flyers, they were people who did not know my linguistic focus.

4.2.4 Participant profile

Over the course of the data collection period, 81 total surveys were completed. As this exceeds the minimum set in advance, 81 was deemed enough to run the statistical analysis for this project.

While the survey was focussed on participants from the UK, it was open for anyone to complete, regardless of country. There were 70 participants from the UK and 11 participants outside the UK. Three participants were from Ireland, three from the USA, two from Canada, two from South Africa, and one participant from Japan.

The participants from the UK were asked to specify which region from the UK they were from. The regions are based on the Government Offices for the English Regions and the inclusion of Scotland, Wales, and Northern Ireland. These regions would be easily identifiable for those within the UK. Based on the 70 participants from the UK, every region was represented in the data apart from Northern Ireland. Table 35 lists the total number of participants for each region, as well as the percentage that region accounts for from the UK data.

Region	Participants	Percentage
Scotland	1	1%
Wales	2	3%
South West	3	4%
South East	13	19%
Greater London	3	4%
East of England	2	3%
West Midlands	5	7%
East Midlands	4	6%
Yorkshire & the Humber	25	36%
North West	5	7%
North East	7	10%

Table 35 Participants from UK regions

In order to account for all forms of gender identity, participants were left an open textbox to describe their gender. Some participants did identify as cis-female or cis-male and others only said female and male; all of these responses were included under the categories of female and male. Table 36 shows all the gender responses from the participants in the study.

Gender	Participants	Percentage
Agender	1	1%
Female	53	65%
Male	21	26%
Trans Male	1	1%
Non-binary	5	6%

Table 36 Participants' gender identity

As with gender, participants were also provided an open textbox to describe their sexual orientation. Responses to sexual orientation for all 81 participants can be seen in Table 37 below. The "other" label refers to responses that could not be categorised, such as people indicating they were unsure of their sexual orientation or describing their sexual orientation as "normal" or "natural". Pansexual was combined with bisexual, as these sexualities are often linked (Flanders *et al.*, 2017) and this combination would allow for easier comparative statistics.

Sexual orientation	Participants	Percentage
Asexual	2	2%
Bisexual	15	19%
Gay	9	11%
Other	4	5%
Queer	4	5%
Straight	47	58%

Table 37 Participants' sexual orientation

Participants were predominantly under 35 years old, but participants over 65 were represented in this data. Table 38 shows the number of participants in each age group, along with the percentage of the total data.

Age	Participants	Percentage
18-24	14	17%
25-34	39	48%
35-44	8	10%
45-54	12	15%
55-64	7	9%
65+	1	1%

Table 38 Participants' age

The final questions participants were asked were if they had any linguistic training, and if so, what the extent of the training was. This was asked to consider if linguists were more sensitive to the shift in F0 than those with no linguistic training. Of the 81 participants, 59 had no linguistic training and 22 participants did have some level of linguistic training. Table 39 outlines the type of linguistic training participants indicated they had had, as well as the percentage this subgroup is of the total data.

Linguistic training	Participants	Percentage
No linguistic degree	59	73%
Current undergrad	4	5%
Completed undergrad	9	11%
Current postgrad	3	4%
Completed postgrad	6	7%

Table 39 Participants' linguistic training

The following section will present the key findings from the survey.

4.3 Results

This section will outline the results from the perception survey described above. Some data will be presented as numerical ratings, but as can be seen in Figure 8, participants were asked to rate based on a written scale of "strongly disagree" to "strongly agree". For the purpose of clarity, Table 40 below has each possible response with the numerical value it is associated with. Both the numerical and textual rating will be used in this results section. The only trait using a different scale is estimated speaker age. These responses were based on age groupings and will be discussed further in the speaker age section below.

Rating	Numeric value
Strongly disagree	1
Disagree	2
Somewhat disagree	3
Neither agree nor disagree	4
Somewhat agree	5
Agree	6
Strongly agree	7

Table 40 Response options with numerical value

4.3.1 Quality Results

The results for each quality will be presented in this section, which will allow for comparisons between the qualities in the correlation section to follow. Some qualities will be given more detailed descriptions, as these were the main focus of the study, while others were used as filler to hide the focus from participants.

Before considering the ratings of each sentence, it is interesting to note that there were interesting patterns in how participants rated each quality more generally. Of the 81 participants, 41 rated every sentence as "neither agree nor disagree" for the quality *homosexual*. An additional participant rated every sentence as "disagree" and for this trait a separate participant rated 38 of the 40 sentences as "disagree". There were only 17 participants out of the total 81 that rated 10 or more sentences as something other than a "neither agree nor disagree" for the quality of *homosexual*.

To put this in perspective, there were five other qualities that participants were asked to rate for each sentence. No single participant rated every sentence as

"neither agree nor disagree" when rating *friendliness* or if a voice was *low pitch*. Only one person rated "Neither agree nor disagree" for every sentence when rating *femininity*, four rated "neither agree nor disagree" every time for *intelligent*, and five rated "Neither agree nor disagree" when rating how *trustworthy* the speaker sounded.

While half the participants stated "neither agree nor disagree" when rating *homosexuality*, the next closest trait to show this much neutrality was *trustworthiness,* and that absolute voting only accounted for 6% of the participants.

This neutrality of voting can also be seen in participants' use of the extremes of the scale voted across the Likert scale. For *friendliness*, different participants rated 11 different sentences at "Strongly disagree" and gave a rating of "Strongly agree" for 38 different sentences. For *intelligence*, 11 different sentences were rated at "Strongly disagree". People voted "Strongly agree" for 23 different sentences. For *trustworthiness* a rating of "Strongly disagree" was given for 6 different sentences from different participants. There was a rating of "strongly agree" for 25 different sentences.

For *femininity*, only two sentences were marked "strongly disagree" by different participants. Interestingly, three participants gave a rating of "strongly agree" for every sentence. For *low pitch*, there were a total of 21 sentences that were rated at "strongly disagree". There was a total of 16 sentences marked at "strongly agree".

Despite this usage of "strongly agree" and "strongly disagree" for the other qualities, this was not the case for ratings of *homosexuality*. Only two participants ever gave a ranking of "strongly agree" when rating *homosexuality*, and eight participants used a rating of "strongly disagree". This shows that participants preferred to stick to the middle ratings when rating *homosexuality* and did not give strong opinions.

To visualise the difference in ratings for *homosexuality* and all the other qualities, a bar chart was created compiling all the votes. Figures 10 through 14 shows all the ratings for the five qualities discussed above, *feminine*, *friendly*, *intelligent*, *low pitch*, and *trustworthy*.
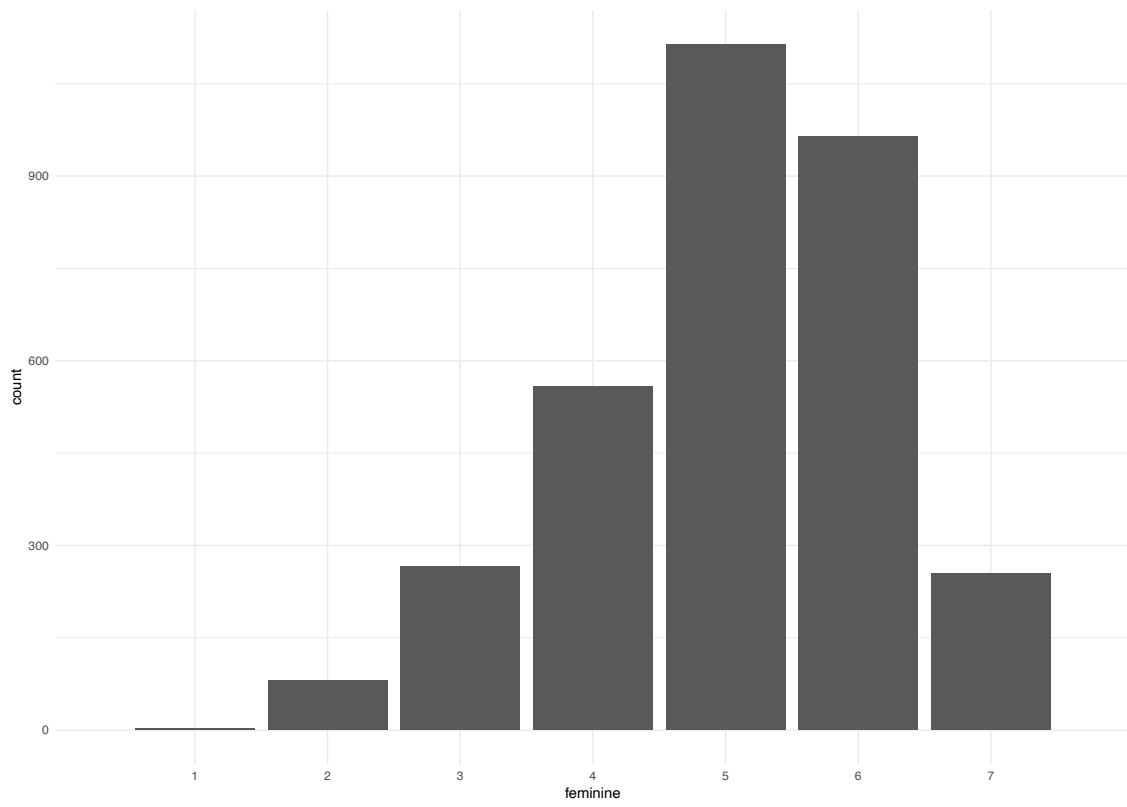
Figure 10 Responses based on the trait *feminine;* 1=strongly disagree and 7=strongly agree
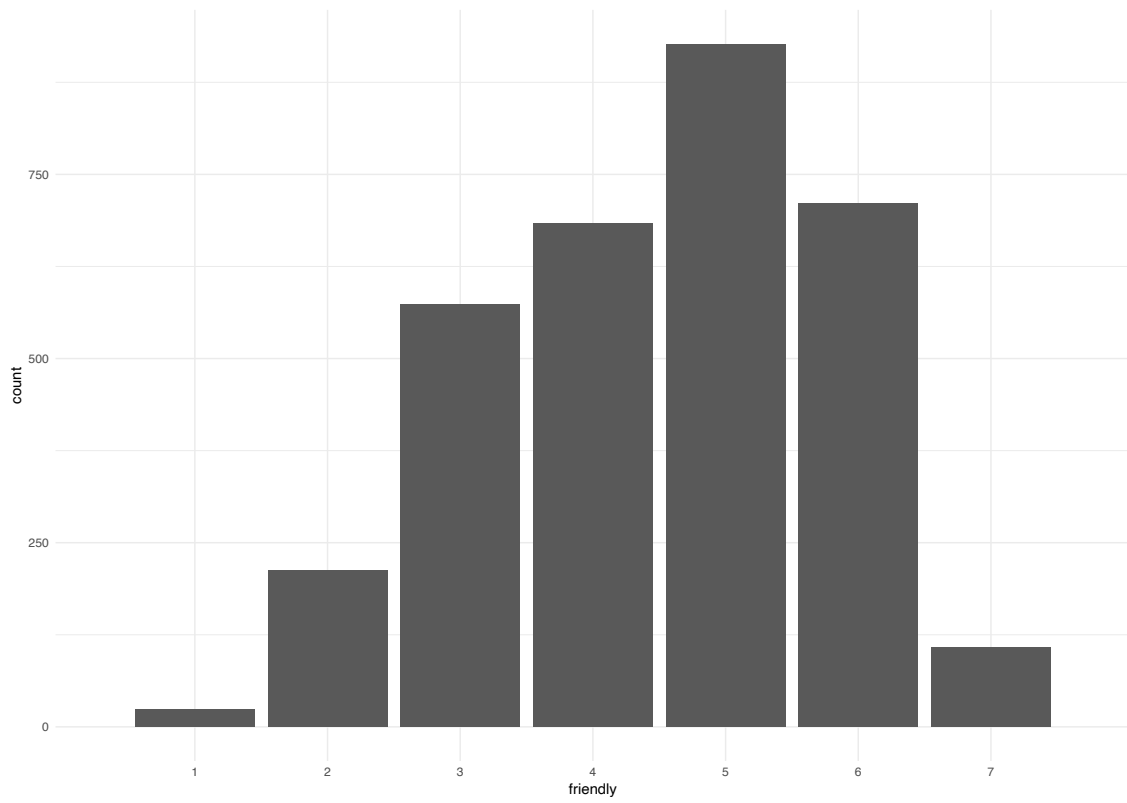
Figure 11 Responses based on the trait *friendly;* 1=strongly disagree and 7=strongly agree

110
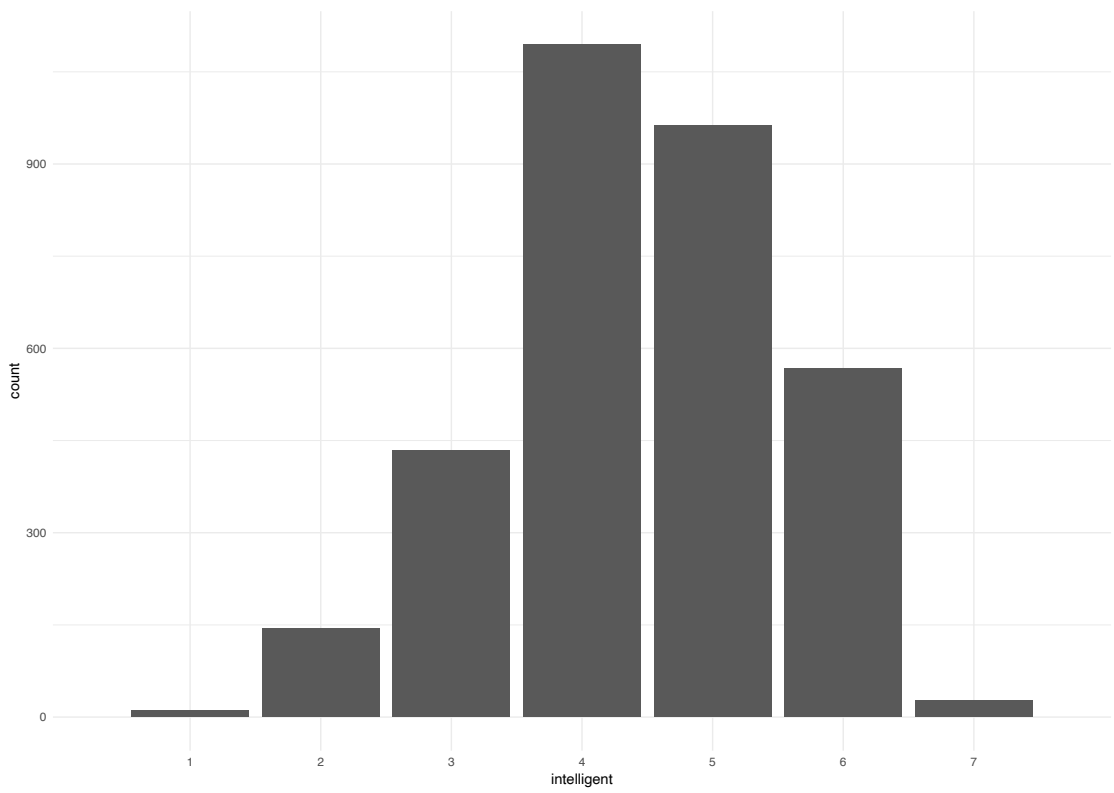
Figure 12 Responses based on the trait *intelligent;* 1=strongly disagree and 7=strongly agree
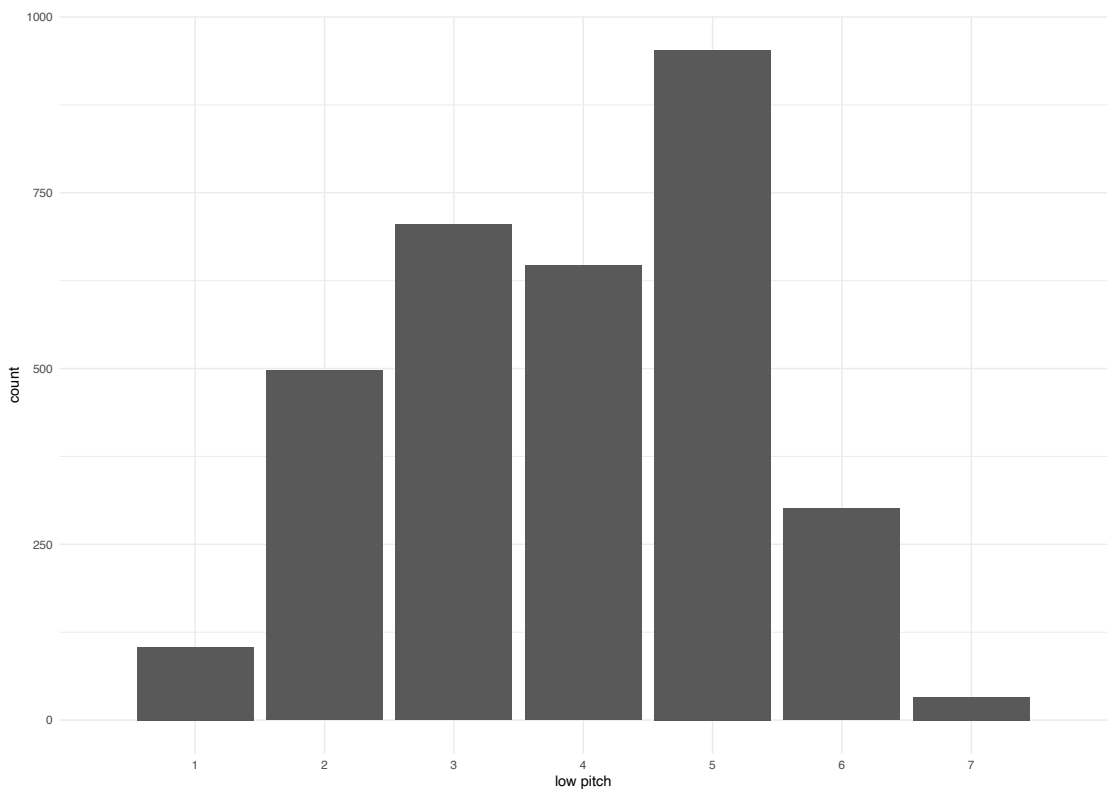
Figure 13 Responses based on the trait *low pitch;* 1=strongly disagree and 7=strongly agree
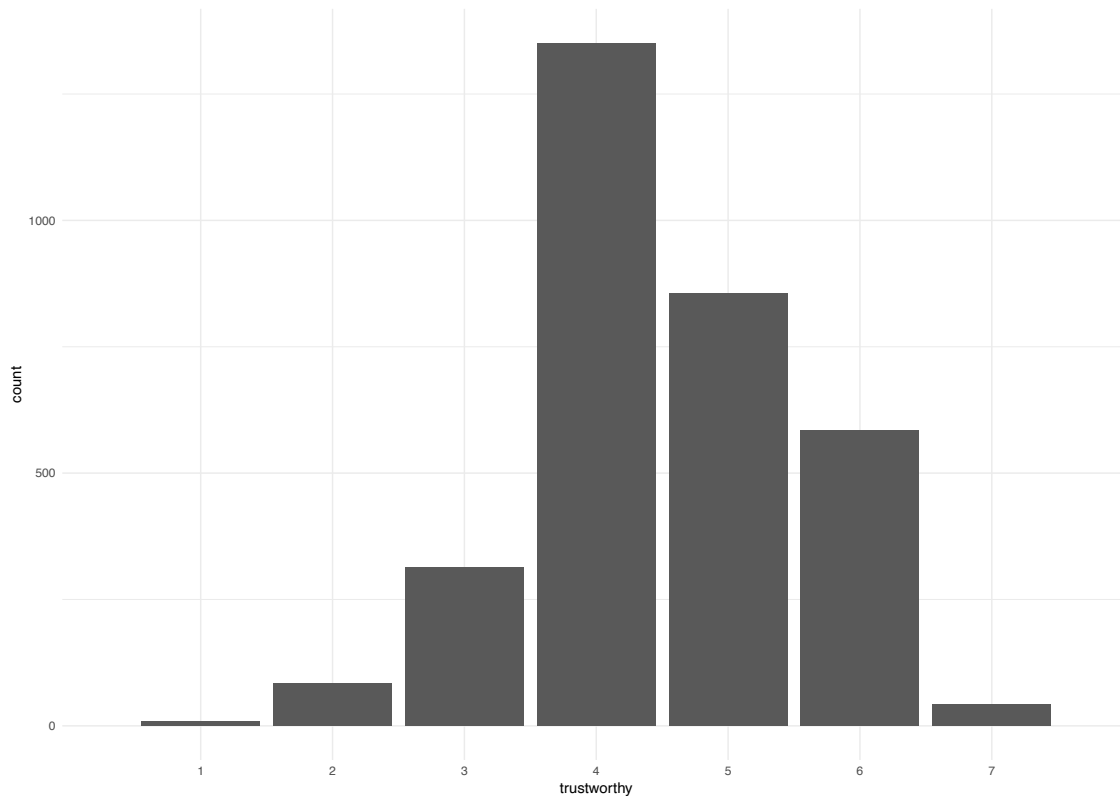
Figure 14 Responses based on the trait *trustworthy;* 1=strongly disagree and 7=strongly agree

Though *intelligent* and *trustworthy* both have the most responses in "neither agree nor disagree", it is clear that there were still votes across all the other levels of the Likert scale. In contrast, in Figure 15 it is clear just how frequently participants rated "neither agree nor disagree" when rating homosexuality.
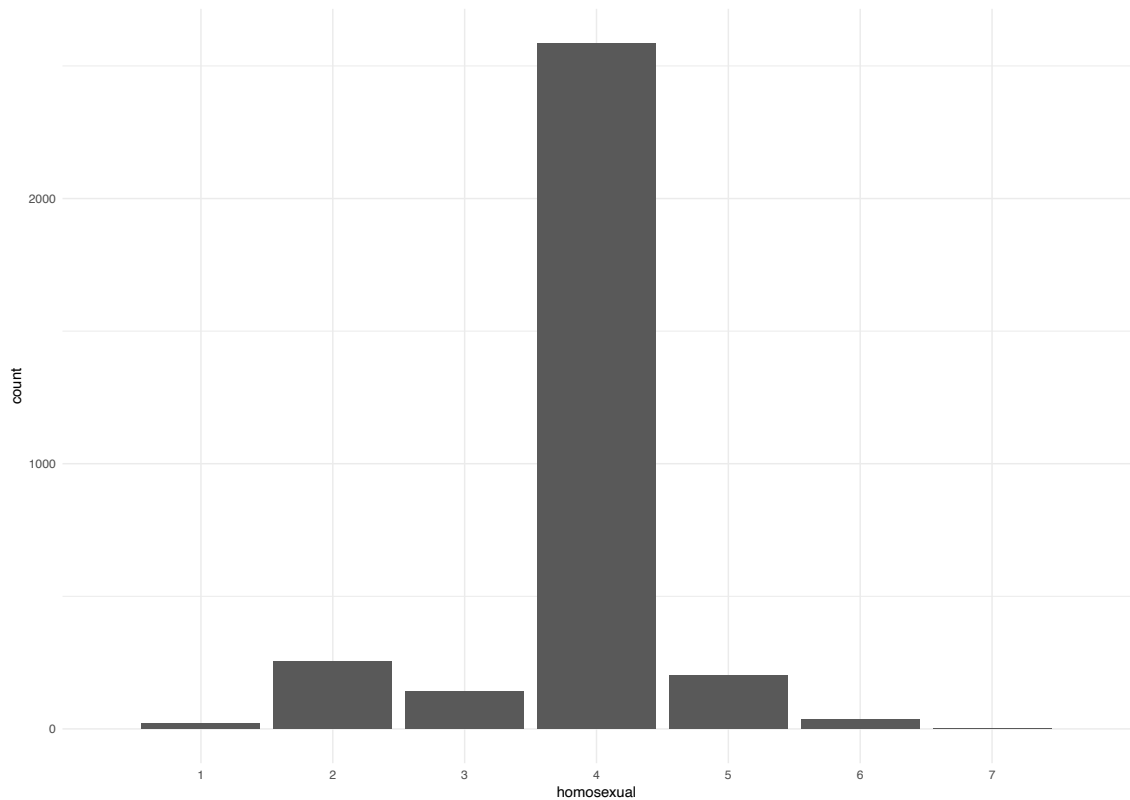
Figure 15 Responses based on the trait *homosexual;* 1=strongly disagree and 7=strongly agree

The discussion below will consider why participants maintained a more neutral voting for homosexuality, despite using broader rating scale for the other qualities.

The following sections will present data on all of the qualities participants were asked to rate. While there were many neutral responses to homosexuality, all the data will be reported, including these absolute votes, because there are still some key findings based on the overall voting trends.

Before conducting any statistical analysis, the data was tested for normality. It was found that the ratings for all of the qualities were significantly different from normal: *feminine* (W = 0.915, p < 0.0001), *friendly* (W = 0.932, p < 0.0001), *homosexual* (W = 0.605, p < 0.0001), *intelligent* (W = 0.921, p < 0.0001), *low pitch* (W = 0.932, p < 0.0001), *speaker age* (W = 0.927, p < 0.0001), and *trustworthy* (W = 0.906, p < 0.0001). As the data was not normally distributed, nonparametric tests were used throughout this results section.

4.3.1.1 Trustworthy

Participants were asked to rate how trustworthy the speaker sounded. This was one of the filler qualities that was included so as to hide the focus on gender and sexual orientation. There is no clear correlation between the level of F0 alteration and the mean ratings of trustworthiness. The data is presented in Table 41 below. The table includes the sentence that was rated, the mean F0 in semitones (based on 100Hz), the mean rating for all the participants, the standard deviation, and the amount the F0 was altered in Hertz for that sentence, if at all. The F0 alteration is colour-coded on a spectrum; red indicates F0 was lowered and the darker the shade, the greater the extent to which its frequency has been lowered. Conversely, green indicates F0 was raised and the darker the shade, the greater extent to which its frequency has been raised. This colour scale allows for clear comparisons in the ratings and will be used for all the qualities in the following section.

Following Table 41, Figure 16 presents a scatterplot of the mean *trustworthy* ratings for every sentence and the mean F0 of that sentence.

| Trustworthy | | | | |
Sentence	F0 (semitones)	Mean	SD	F0 Alteration
2	5.3	3.58	1.34	-50
29	14.8	3.86	1.17	50
30	8.1	4.23	1.06	-30
13	8.6	4.31	0.96	-30
36	6.3	4.31	1.09	-50
10	13.3	4.35	0.99	20
34	13.2	4.35	0.99	20
28	8.9	4.37	1.04	-20
8	12.2	4.40	0.94	-20
40	10.9	4.40	1.01	0
24	16.4	4.41	0.95	40
4	11	4.42	1.00	-30
12	11.1	4.42	1.02	0
18	5.3	4.43	1.20	-50
25	13.1	4.43	1.00	20
32	8	4.47	0.98	-40
23	11.8	4.48	0.94	-20
9	15.6	4.49	0.98	30
31	14	4.49	1.00	30
3	7.4	4.51	1.03	-40
5	14.4	4.53	1.13	40
17	13.4	4.54	1.01	20
7	7.9	4.56	0.92	-40
26	7.1	4.57	1.02	-50
33	17.1	4.59	1.01	50
37	6.6	4.60	0.89	-40
38	13.9	4.60	0.86	30
16	14	4.62	1.01	30
14	15.5	4.63	1.05	50
21	9.1	4.64	0.97	-40
19	7.7	4.65	0.96	-30
39	9.1	4.65	0.98	-30
35	16.8	4.69	0.88	40
11	14.6	4.70	1.02	40
15	9	4.74	0.95	-20
27	14.3	4.74	0.89	30
20	15.8	4.78	0.97	50
1	12.1	4.90	1.04	0
6	16.5	4.90	1.09	50
22	14.1	4.98	0.94	0

Table 41 *Trustworthy* ratings, listed lowest mean rating to highest mean rating
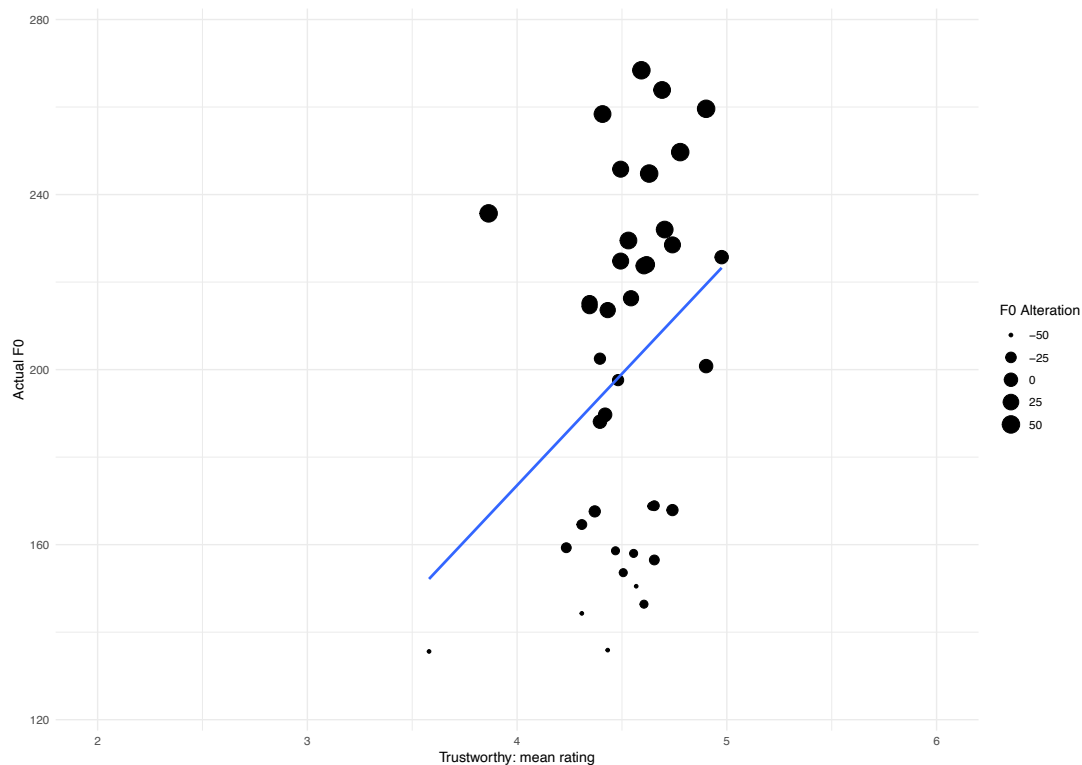
Figure 16 Trustworthy mean rating with actual F0

A Spearman's correlation test suggests that mean *trustworthy* ratings are not significantly correlated with the amount of F0 alteration ($r = 0.237$, $p = 0.141$). These results may imply that speakers did not have strong preconceived notions about what a trustworthy voice would sound like, not only based on this lack of correlation, but also because this quality was the second highest to have unanimous agreement across the listener sample in the option of "neither agree nor disagree".

4.3.1.2 Intelligent

A second quality that was asked about as a filler in the survey was intelligence. However, unlike trustworthiness, there appears to be more of a pattern with respect to how intelligent a female speaker sounds, based on her mean F0. Table 42 presents the ratings for *intelligent*, presented in the same way as Table 39. Following Table 42, Figure 17 presents a scatterplot of the mean *intelligent* ratings for every sentence and the mean F0 of that sentence.

Intelligence				
Sentence	F0 (semitones)	Mean	SD	F0 Alteration
12	11.1	3.98	1.21	0
14	15.5	4.06	1.21	+50
3	7.4	4.14	1.23	-40
11	14.6	4.14	1.06	+40
31	14	4.17	1.18	+30
5	14.4	4.19	1.10	+40
40	10.9	4.23	1.12	0
8	12.2	4.25	1.08	-20
16	14	4.26	1.15	+30
13	8.6	4.27	1.11	-30
20	15.8	4.28	1.24	+50
9	15.6	4.30	1.13	+30
27	14.3	4.32	1.15	+30
30	8.1	4.32	1.15	-30
39	9.1	4.36	1.25	-30
34	13.2	4.38	1.12	+20
24	16.4	4.40	1.06	+40
29	14.8	4.41	1.09	+50
33	17.1	4.41	1.16	+50
25	13.1	4.42	0.95	+20
4	11	4.43	1.01	-30
10	13.3	4.43	1.15	+20
23	11.8	4.46	0.99	-20
35	16.8	4.46	1.07	+40
17	13.4	4.49	1.03	+20
38	13.9	4.51	1.04	+30
6	16.5	4.52	1.15	+50
7	7.9	4.52	1.11	-40
18	5.3	4.53	1.17	-50
15	9	4.54	1.00	-20
32	8	4.56	1.05	-40
36	6.3	4.57	1.09	-50
2	5.3	4.60	1.07	-50
28	8.9	4.60	1.09	-20
37	6.6	4.70	0.94	-40
21	9.1	4.72	1.05	-40
19	7.7	4.78	0.89	-30
1	12.1	4.86	0.92	0
26	7.1	4.98	0.88	-50
22	14.1	5.04	0.89	0

Table 42 *Intelligent* ratings, listed lowest mean rating to highest mean rating
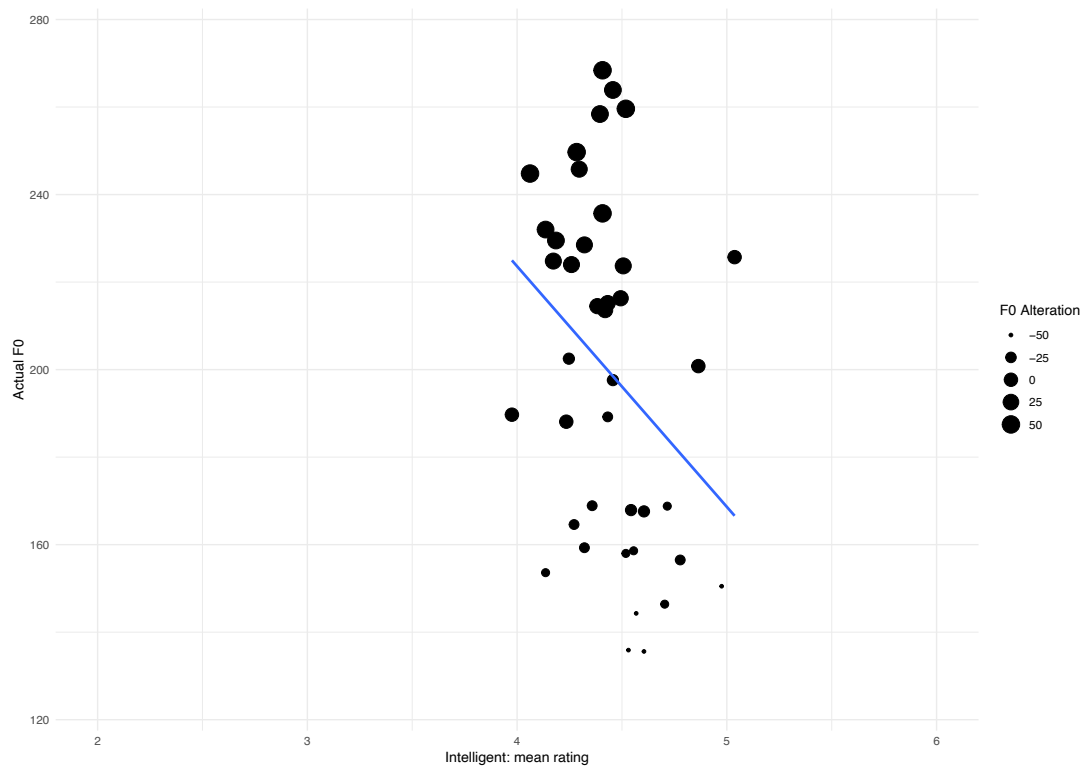
Figure 17 Intelligent mean rating with actual F0

Here we see clear favouring of sentences that had the F0 digitally altered to be lower being rated as more *intelligent*. This correlation based on mean rating is confirmed to be significant when subjected to a Spearman's correlation test (r = -0.474, p = 0.002).

4.3.1.3 Friendly

The final quality that acted as a filler in the survey was *friendly*. As with intelligence, there are some indications that friendliness is rated based on the speaker's F0. Following Table 43, Figure 18 presents a scatterplot of the mean *friendly* ratings for every sentence and the mean F0 of that sentence.

	Friendly			
Sentence	F0 (semitones)	Mean	SD	F0 Alteration
36	6.3	3.40	1.41	-50
2	5.3	3.42	1.42	-50
10	13.3	3.70	1.42	+20
13	8.6	3.91	1.21	-30
29	14.8	4.07	1.32	+50
37	6.6	4.11	1.28	-40
26	7.1	4.12	1.41	-50
18	5.3	4.14	1.46	-50
32	8	4.16	1.26	-40
34	13.2	4.16	1.16	+20
28	8.9	4.20	1.35	-20
19	7.7	4.22	1.38	-30
30	8.1	4.27	1.16	-30
7	7.9	4.28	1.28	-40
8	12.2	4.38	1.37	-20
12	11.1	4.41	1.38	0
4	11	4.43	1.22	-30
25	13.1	4.48	1.21	+20
9	15.6	4.49	1.38	+30
23	11.8	4.63	1.05	-20
17	13.4	4.64	1.23	+20
21	9.1	4.64	1.13	-40
40	10.9	4.65	1.07	0
14	15.5	4.69	1.30	+50
31	14	4.69	1.13	+30
1	12.1	4.70	1.26	0
27	14.3	4.70	1.25	+30
5	14.4	4.74	1.34	+40
15	9	4.74	1.17	-20
3	7.4	4.77	1.38	-40
39	9.1	4.78	1.19	-30
33	17.1	4.80	1.09	+50
16	14	4.83	1.27	+30
24	16.4	4.88	1.22	+40
38	13.9	4.89	1.07	+30
35	16.8	4.90	1.17	+40
11	14.6	4.93	1.16	+40
22	14.1	5.11	1.00	0
20	15.8	5.32	1.03	+50
6	16.5	5.37	1.04	+50

Table 43 *Friendly* ratings, listed lowest mean rating to highest mean rating
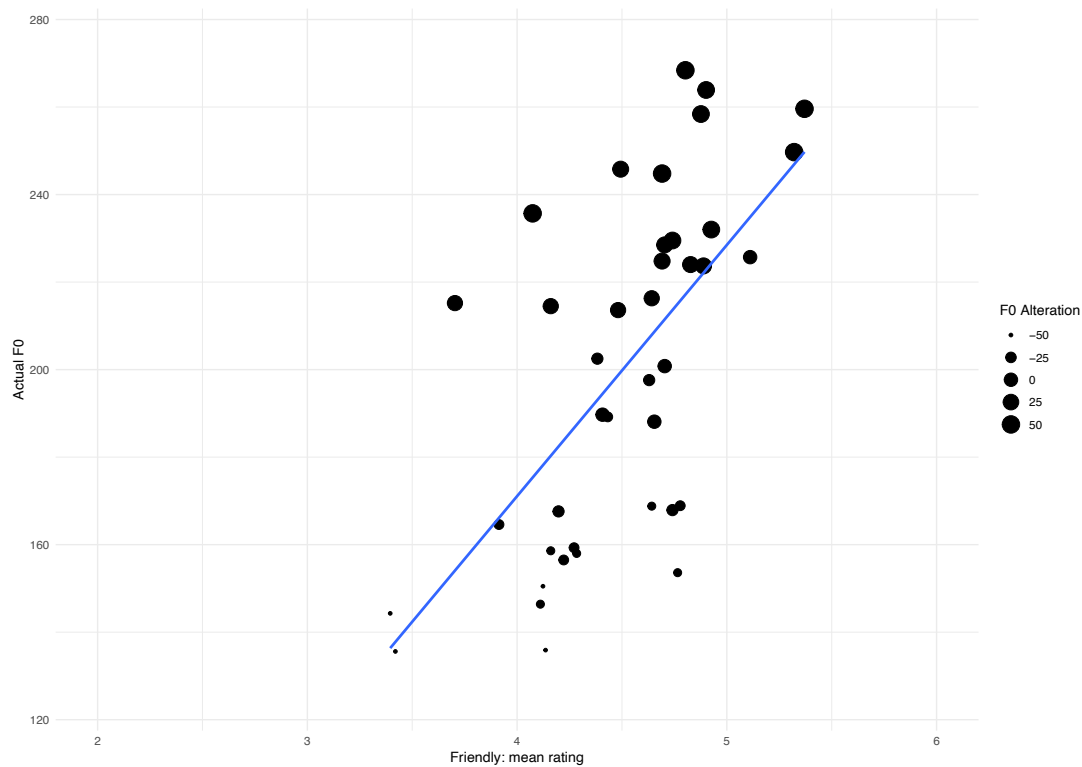
Figure 18 Friendly mean rating with actual F0

A Spearman's correlation test indicates that there is a positive correlation between F0 alteration and mean friendliness ratings ($r = 0.61$, $p < 0.0001$). This data suggests that female speakers with a higher mean F0 are perceived as friendlier than female speakers with a lower mean F0.

4.3.1.4 Speaker Age

Speaker age was included in this study in order to test the hypothesis that a voice with a lower F0 would be perceived as gay or older, or possibly both. As with all the other categories, a numeric value has been given to each rating, but these ratings are based on age groups instead of level of agreement. Table 44 provides a key to the age ratings and the numeric values they received.

Rating	Numeric value
Under 18	1
18-24	2
25-34	3
35-44	4
45-54	5
55-64	6
65 and older	7

Table 44 Age response options with numeric values

Participants' ratings for the age of the speaker are presented below in Table 45. Following Table 45, Figure 19 presents a scatterplot of the mean speaker age ratings for every sentence and the mean F0 of that sentence.

	Speaker Age			
Sentence	F0 (semitones)	Mean	SD	F0 Alteration
29	14.8	2.46	0.81	+50
14	15.5	2.48	0.71	+50
20	15.8	2.54	0.65	+50
5	14.4	2.58	0.83	+40
31	14	2.64	0.81	+30
16	14	2.74	0.70	+30
11	14.6	2.81	0.91	+40
35	16.8	2.84	1.09	+40
27	14.3	2.88	0.87	+30
6	16.5	2.90	0.94	+50
24	16.4	2.90	0.72	+40
38	13.9	3.04	1.04	+30
40	10.9	3.20	0.95	0
10	13.3	3.25	0.90	+20
33	17.1	3.31	0.85	+50
34	13.2	3.35	0.87	+20
12	11.1	3.37	0.98	0
17	13.4	3.48	0.98	+20
25	13.1	3.49	1.09	+20
9	15.6	3.63	1.02	+30
30	8.1	3.75	0.81	-30
15	9	3.78	0.95	-20
3	7.4	3.80	1.04	-40
39	9.1	3.96	0.93	-30
4	11	3.98	0.77	-30
22	14.1	4.01	0.94	0
21	9.1	4.05	0.88	-40
1	12.1	4.06	0.94	0
23	11.8	4.15	0.91	-20
2	5.3	4.20	0.97	-50
7	7.9	4.23	0.76	-40
8	12.2	4.33	1.05	-20
13	8.6	4.33	0.81	-30
19	7.7	4.35	0.84	-30
28	8.9	4.40	0.93	-20
37	6.6	4.51	0.90	-40
36	6.3	4.58	0.83	-50
18	5.3	4.63	0.93	-50
32	8	4.91	0.99	-40
26	7.1	4.93	0.86	-50

Table 45 *Speaker age* ratings, listed lowest mean rating to highest mean rating
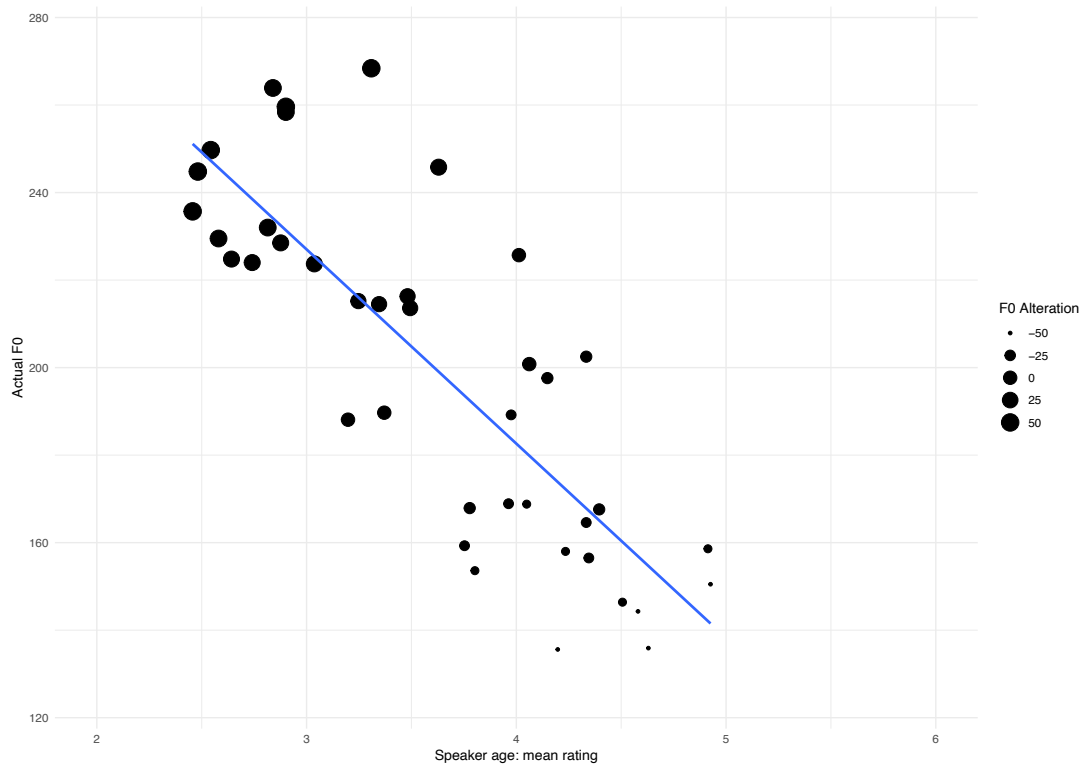
Figure 19 Speaker age mean rating with actual F0

It is clear that participants were sensitive to the pitch alterations and felt that a lower pitch was associated with an older speaker. When the average F0 of a sentence was lowered by 40 to 50Hz, the mean rating for the speaker's age group was 35-44 and moving towards the 45-54 age category. A Spearman's correlation test confirms that there is a significant negative correlation between mean age ratings and F0 alteration (r = -0.888, p < 0.0001).

4.3.1.5 Low Pitch

Speaker age is not the only quality that showed how sensitive participants were to the speaker's F0. Participants were directly asked if they believed the speaker had a low-pitched voice in order to assess how aware they were of this quality. Table 46 below shows just how sensitive and accurate participants' pitch perception appear to be. Following Table 46, Figure 20 presents a scatterplot of the mean low pitch ratings for every sentence and the mean F0 of that sentence.

		Low Pitch		
Sentence	F0 (semitones)	Mean	SD	F0 Alteration
6	16.5	2.62	1.12	+50
20	15.8	2.74	1.18	+50
33	17.1	2.78	1.06	+50
14	15.5	2.88	1.17	+50
24	16.4	2.98	1.20	+40
29	14.8	3.05	1.16	+50
35	16.8	3.06	1.17	+40
38	13.9	3.12	1.22	+30
16	14	3.14	1.09	+30
9	15.6	3.15	1.06	+30
5	14.4	3.16	1.24	+40
31	14	3.16	1.20	+30
27	14.3	3.31	1.21	+30
17	13.4	3.37	1.15	+20
11	14.6	3.43	1.21	+40
22	14.1	3.57	1.25	0
25	13.1	3.65	1.20	+20
34	13.2	3.70	1.13	+20
40	10.9	3.72	1.11	0
1	12.1	3.86	1.29	0
10	13.3	3.88	1.21	+20
12	11.1	3.91	1.21	0
4	11	4.10	1.17	-30
23	11.8	4.20	1.10	-20
3	7.4	4.26	1.23	-40
39	9.1	4.32	1.28	-30
8	12.2	4.35	1.18	-20
15	9	4.41	1.18	-20
28	8.9	4.49	1.10	-20
21	9.1	4.57	1.02	-40
13	8.6	4.62	0.97	-30
19	7.7	4.62	1.07	-30
30	8.1	4.62	1.07	-30
2	5.3	4.68	1.26	-50
7	7.9	4.77	1.06	-40
37	6.6	4.85	0.92	-40
32	8	5.00	0.97	-40
26	7.1	5.11	1.10	-50
18	5.3	5.19	0.95	-50
36	6.3	5.22	1.05	-50

Table 46 *Low pitch* rating, listed lowest mean rating to highest mean rating

124
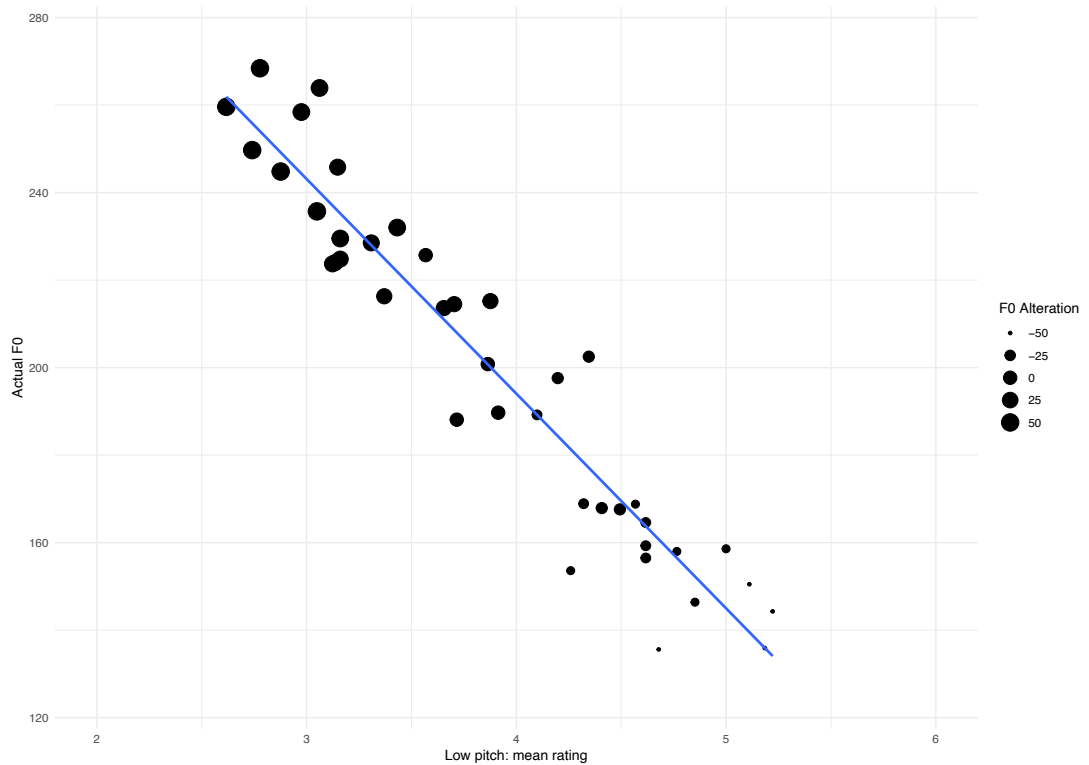
Figure 20 Low pitch mean rating with actual F0

The table demonstrates that participants were very sensitive to small variations in pitch and accurately rated sentences based on the degree of alteration. A Spearman's correlation test indicates there is a significant negative correlation between mean pitch rating and F0 alteration ($r = -0.96$, $p < 0.0001$). While there are a few exceptions, participants were even aware of the 10Hz difference, as both the highest and lowest ratings have alterations of 50Hz.

4.3.1.6 Feminine

Another quality that proved to be strongly impacted by F0 alteration was perceived femininity. Table 47 reports the results of femininity ratings based on F0 alteration. Following Table 47, Figure 21 presents a scatterplot of the mean *feminine* ratings for every sentence and the mean F0 of that sentence.

	Feminine			
Sentence	F0 (semitone)	Mean	SD	F0 Alteration
36	6.3	4.30	1.36	-50
13	8.6	4.53	1.33	-30
26	7.1	4.56	1.32	-50
30	8.1	4.64	1.25	-30
2	5.3	4.67	1.33	-50
37	6.6	4.67	1.28	-40
7	7.9	4.72	1.28	-40
18	5.3	4.74	1.23	-50
28	8.9	4.74	1.22	-20
12	11.1	4.75	1.17	0
8	12.2	4.77	1.26	-20
23	11.8	4.77	1.13	-20
32	8	4.80	1.19	-40
19	7.7	4.81	1.23	-30
21	9.1	4.83	1.21	-40
39	9.1	4.88	1.16	-30
9	15.6	4.93	1.12	+30
3	7.4	4.96	1.22	-40
15	9	4.96	1.11	-20
10	13.3	5.00	1.12	+20
25	13.1	5.04	1.02	+20
34	13.2	5.05	1.02	+20
4	11	5.07	1.06	-30
40	10.9	5.09	1.07	0
17	13.4	5.19	1.05	+20
38	13.9	5.20	0.94	+30
22	14.1	5.21	1.03	0
29	14.8	5.26	1.01	+50
27	14.3	5.27	1.04	+30
33	17.1	5.33	0.96	+50
35	16.8	5.33	1.01	+40
31	14	5.37	1.01	+30
11	14.6	5.38	0.92	+40
24	16.4	5.40	0.90	+40
5	14.4	5.48	1.04	+40
16	14	5.48	1.04	+30
20	15.8	5.54	0.95	+50
14	15.5	5.57	0.92	+50
1	12.1	5.68	0.97	0
6	16.5	5.69	0.82	+50

Table 47 *Feminine* rating, listed lowest mean rating to highest mean rating
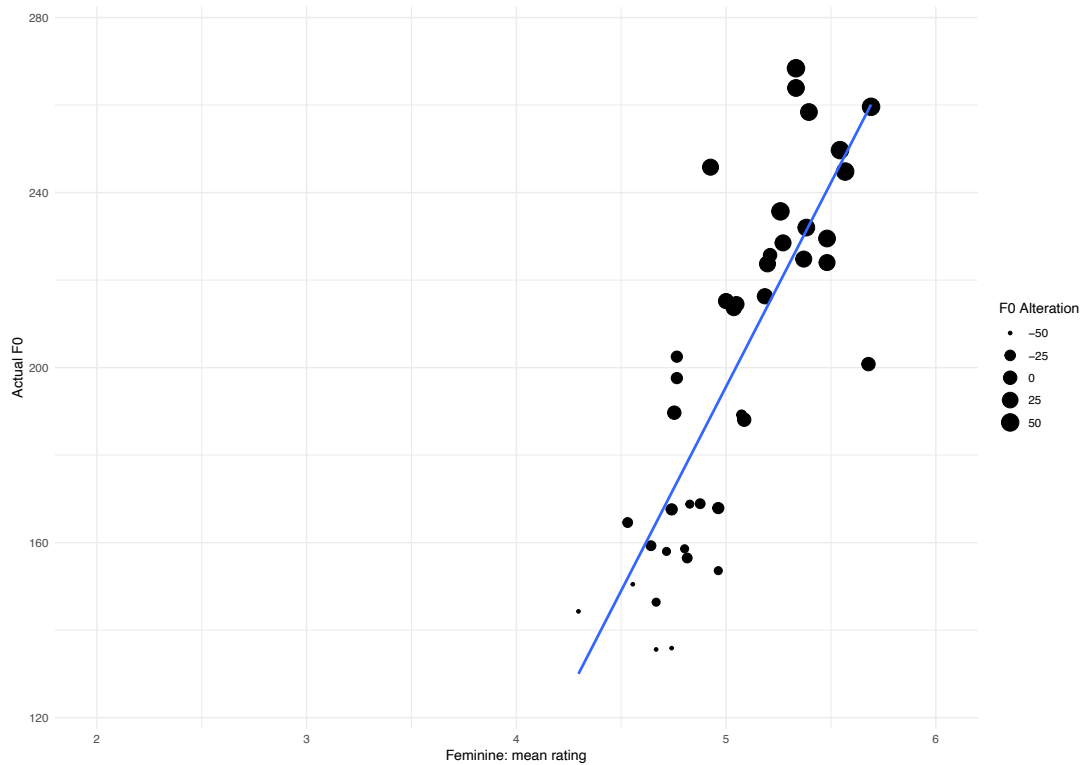
Figure 21 Feminine mean rating with actual F0

Here it is clear that F0 alteration influenced perceptions of femininity and a Spearman's correlation test confirms there is a positive correlation between perceived mean femininity and F0 alterations ($r = 0.846$, $p < 0.0001$).

4.3.1.7 Homosexual

The final quality to consider, and the focus of the current study, is homosexuality. While it was outlined above that participants generally rated "neither agree nor disagree" when responding to whether a sentence sounded *homosexual*, patterns still emerge in the data. Table 48 shows all the responses to homosexuality. Following Table 48, Figure 22 presents a scatterplot of the mean *homosexual* ratings for every sentence and the mean F0 of that sentence.

	Homosexual			
Sentence	F0 (semitone)	Mean	SD	F0 Alteration
1	12.1	3.67	0.94	0
20	15.8	3.68	0.69	+50
14	15.5	3.69	0.82	+50
6	16.5	3.72	0.71	+50
27	14.3	3.72	0.73	+30
11	14.6	3.74	0.69	+40
17	13.4	3.74	0.74	+20
16	14	3.75	0.70	+30
33	17.1	3.75	0.72	+50
38	13.9	3.77	0.87	+30
29	14.8	3.78	0.65	+50
35	16.8	3.80	0.70	+40
9	15.6	3.83	0.70	+30
24	16.4	3.83	0.70	+40
31	14	3.83	0.74	+30
34	13.2	3.83	0.85	+20
32	8	3.84	0.80	-40
5	14.4	3.85	0.74	+40
10	13.3	3.85	0.67	+20
4	11	3.86	0.72	-30
40	10.9	3.88	0.64	0
26	7.1	3.89	0.63	-50
28	8.9	3.90	0.72	-20
12	11.1	3.91	0.60	0
22	14.1	3.91	0.67	0
37	6.6	3.91	0.81	-40
39	9.1	3.91	0.67	-30
13	8.6	3.93	0.61	-30
8	12.2	3.94	0.75	-20
19	7.7	3.94	0.75	-30
15	9	3.95	0.61	-20
3	7.4	3.96	0.75	-40
18	5.3	3.96	0.68	-50
7	7.9	3.98	0.61	-40
23	11.8	3.98	0.57	-20
36	6.3	3.99	0.70	-50
21	9.1	4.00	0.63	-40
30	8.9	4.02	0.71	-30
25	13.1	4.04	0.58	+20
2	5.3	4.09	0.67	-50

Table 48 *Homosexual* rating, listed lowest mean rating to highest mean rating
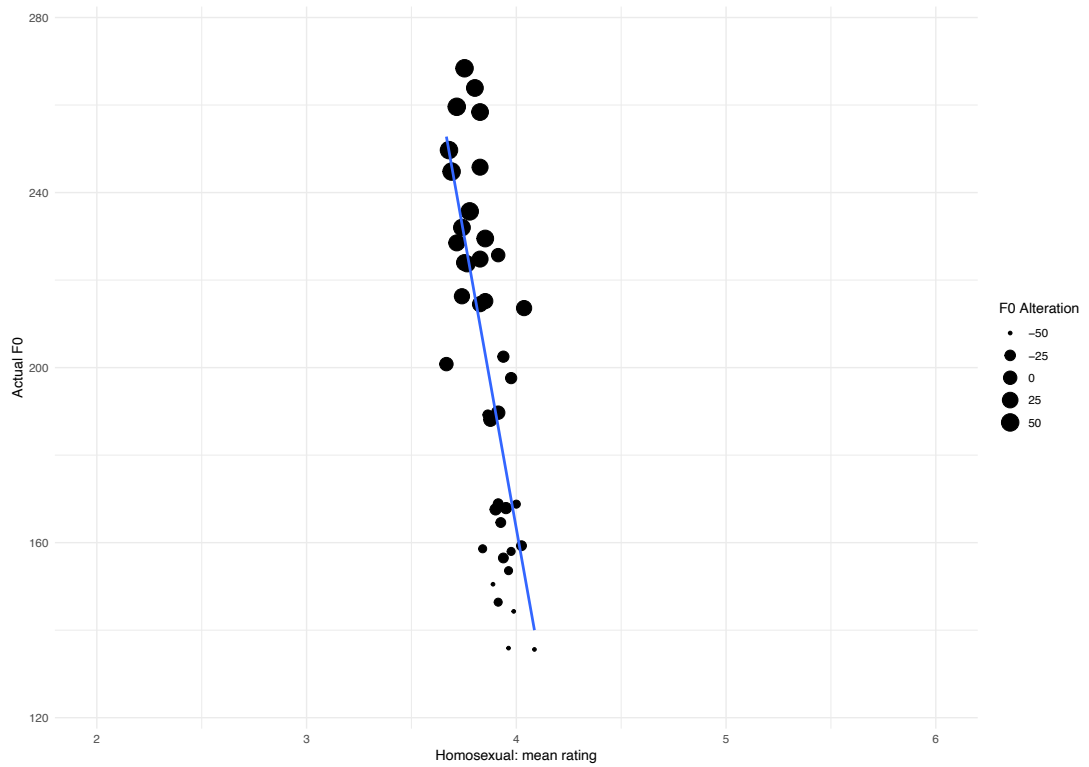
Figure 22 Homosexual mean rating with actual F0

There is a clear tendency for listeners to rate sentences with decreased F0 as sounding more homosexual than sentences with increased F0. A Spearman's correlation test indicates that there is a significant negative correlation between mean *homosexual* ratings and F0 alteration (r = -0.768, p < 0.0001).

In order to compare the mean ratings for each quality, Figure 23 presents all the previous scatter plots side by side.
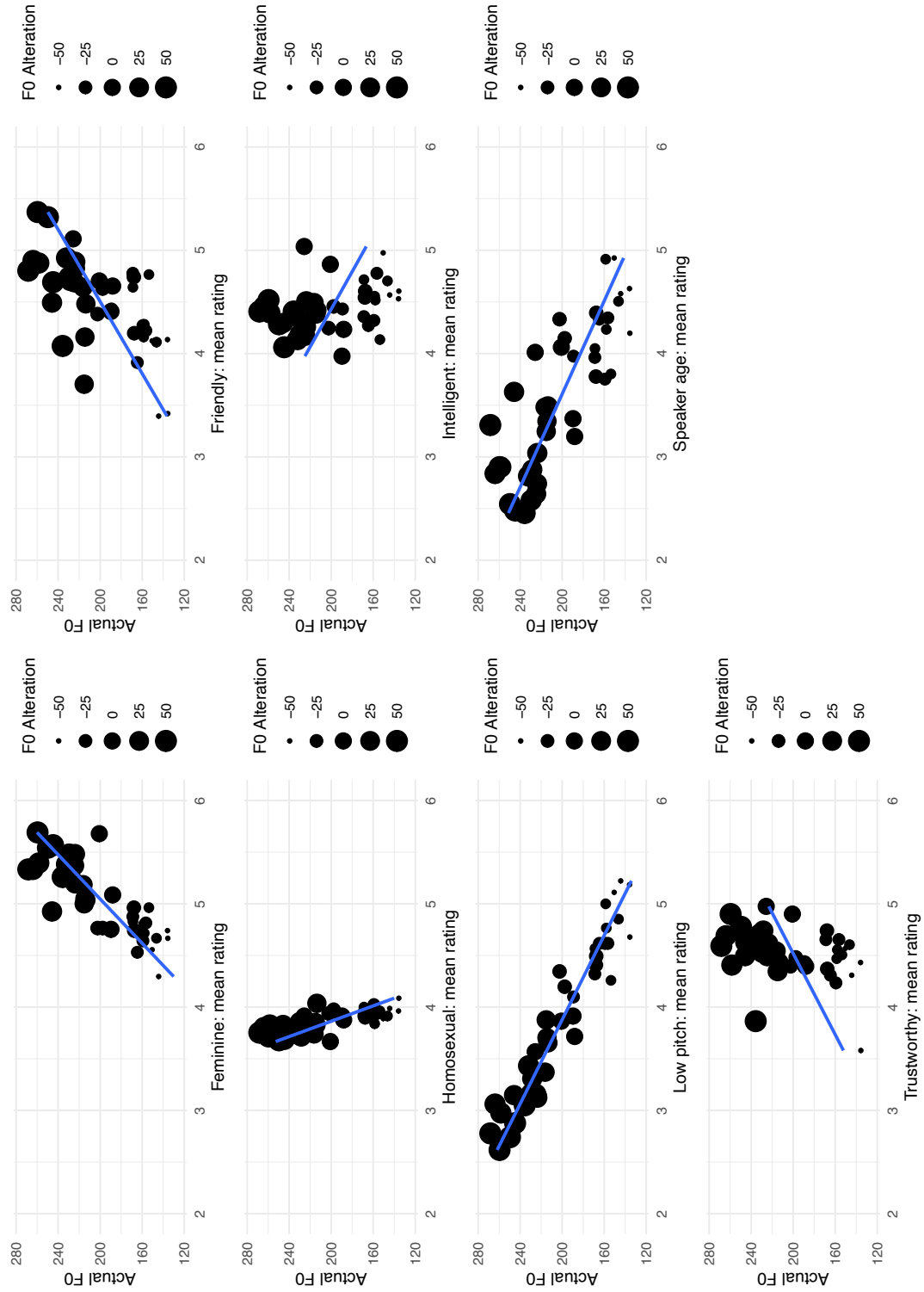
Figure 23 All quality mean ratings with actual F0

4.3.2 Group comparisons

Having considered how the entire participant pool voted, it is important to understand how different groups viewed the data.

4.3.2.1 Sexual orientation comparisons

This section concerns itself with investigating whether listener sexual orientation had an effect on how qualities were rated, and in particular how homosexuality and femininity were rated. As addressed above, participants were initially combined into larger groups in order to allow for an initial comparison. All participants that identified as straight are in a group labelled "straight", participants that identified as gay, queer, bisexual, or asexual are in a group labelled "LGBTQ", and finally participants whose sexual orientation was marked as "other" were excluded from this portion of the results. After the results are presented based on the larger grouping, results will then be presented for the more precise sexual orientation identifications.

The first quality that will be considered across sexual orientation lines is *feminine*. A two-tailed Mann-Whitney U-test suggests that the feminine ratings between the LGBTQ group and the straight group are not significantly different ($U = 1111900$, $p = 0.486$). Figure 24 shows the general overlap between the groups of participants based on sexual orientation.
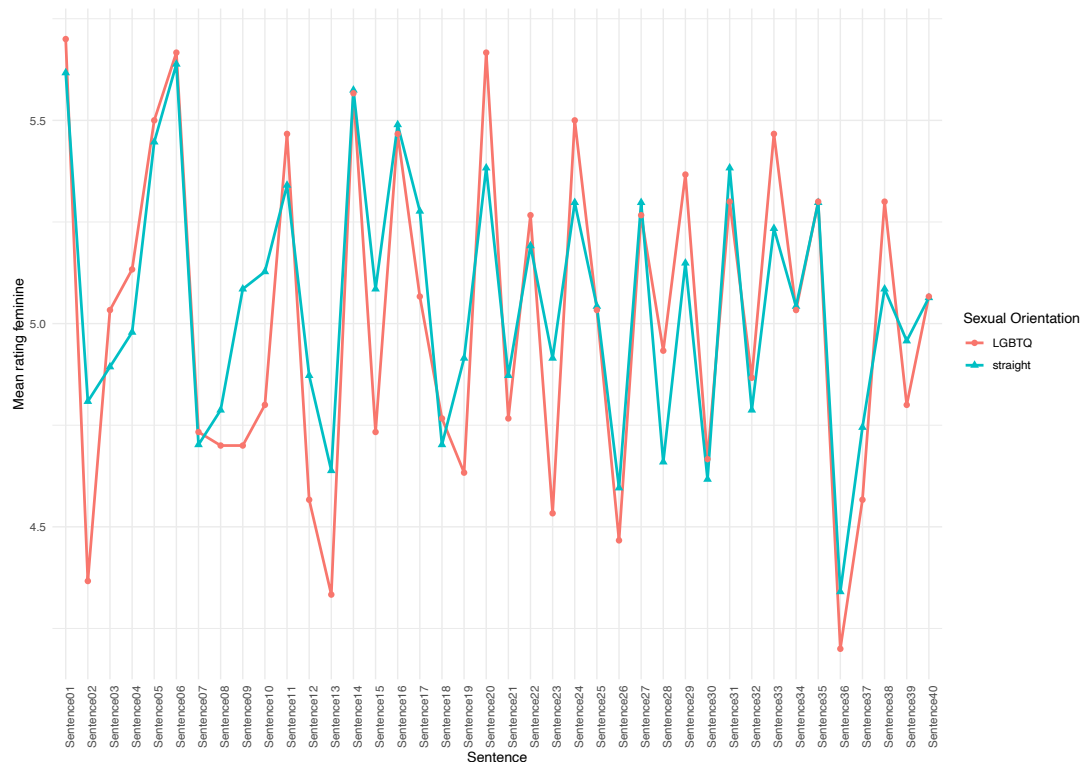
Figure 24 *Feminine* mean rating grouped by sexual orientation

This is further confirmed when considering the sentences rated highest for femininity for both groups of speakers, as shown in in Table 49.

LGBTQ (n=30)			Straight (n=47)		
Sentence	**Mean**	**F0 Alteration**	**Sentence**	**Mean**	**F0 Alteration**
1	**5.70**	**0**	**6**	**5.64**	**+50**
6	**5.67**	**+50**	**1**	**5.62**	**0**
20	**5.67**	**+50**	**14**	**5.57**	**+50**
14	**5.57**	**+50**	**16**	**5.49**	**+30**
5	**5.50**	**+40**	**5**	**5.45**	**+40**
24	**5.50**	**+40**	**20**	**5.38**	**+30**
11	**5.47**	**+50**	31	5.38	+50
16	**5.47**	**+30**	**11**	**5.34**	**+40**
33	5.47	+40	**24**	**5.30**	**+40**
29	5.37	+50	27	5.30	+40
			35	5.30	+30

Table 49 Highest *feminine* ratings by LGBTQ and straight speakers; sentences that appear on both lists emboldened

The similarities between the groups can also be seen in the sentences with the lowest feminine ratings, as seen in Table 50.

LGBTQ (n=30)			Straight (n=47)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
36	**4.20**	**-50**	**36**	**4.34**	**-50**
13	**4.33**	**-30**	**26**	**4.60**	**-50**
2	4.37	-50	**30**	**4.62**	**-30**
26	**4.47**	**-50**	**13**	**4.64**	**-30**
23	4.53	-20	28	4.66	-28
12	4.57	0	7	4.70	-40
37	**4.57**	**-40**	18	4.70	-50
19	4.63	-30	**37**	**4.74**	**-40**
30	**4.67**	**-30**	**8**	**4.79**	**-20**
8	**4.70**	**-20**	32	4.79	-40
9	4.70	30			

Table 50 Lowest *feminine* ratings by LGBTQ and straight speakers; sentences that appear on both lists emboldened

There are many sentences that appear on both groups' highest- and lowest-rated 10 sentences. The similarities between the two groups can also be seen in the rating range for both sexual orientation groups. The sentence given the lowest average *femininity* rating by the LGBTQ group yielded a score of 4.2, while that ranked highest had an average score of 5.7, giving the LGBTQ a mean *feminine* rating range of 1.5. The sentence given lowest mean *feminine* rating for the straight group yielded a score of 4.34, and the highest rated had score of 5.64, giving the straight group a mean *feminine* rating range of 1.3. These ranges demonstrate just how close the groups are in terms of their ratings for *femininity*.

It is evident that both groups rated sentences with an increased F0 as more *feminine* and sentences with a decreased F0 as less *feminine*. Even sentences that do not appear on both groups' highest- and lowest-rated lists follow this same pattern. In fact, there is not a single sentence that is rated significantly differently by these two groups; a table of these results can be seen in Table 51. It is clear that LGBTQ and straight participants did not perceive femininity differently in this study.

Sentence	U	p	Sentence	U	p
1	781	0.401	21	683	0.817
2	578.5	0.177	22	766.5	0.499
3	727.5	0.809	23	591.5	0.22
4	787	0.369	24	765	0.508
5	711.5	0.947	25	689	0.865
6	707.5	0.982	26	653	0.579
7	708.5	0.974	27	673	0.727
8	677.5	0.77	28	793	0.345
9	574	0.157	29	819.5	0.209
10	569.5	0.134	30	738	0.727
11	756	0.574	31	697.5	0.938
12	597	0.239	32	734	0.758
13	610.5	0.314	33	773.5	0.454
14	682.5	0.808	34	719.5	0.878
15	603	0.261	35	701.5	0.974
16	676.5	0.755	36	658.5	0.622
17	635	0.443	37	668	0.695
18	734	0.758	38	789.5	0.351
19	606	0.286	39	657.5	0.609
20	812	0.237	40	698	0.943

Table 51 Mann-Whitney U-test for each sentence based on *feminine* ratings between LGBTQ and straight participants

Having considered whether there were different ratings assigned by these two broad sexual orientation-based groups, it is also important to investigate whether there are rating differences between the narrower sexual identity categories. A Kruskal-Wallis H-test suggests that there are significant *feminine* rating differences between sexual identity groups (Chi square = 57.37, p <0.000039, df = 4). A Mann-Whitney U-test was then run on each possible combination of sexual orientation in order to locate where the difference occurs; see Table 52 for the results.

Sexual orientation comparisons	U	p	Sexual orientation comparisons	U	p
Asexual ~ bisexual	30820	<0.0001	Bisexual ~ queer	36843	<0.0001
Asexual ~ gay	18732	<0.0001	Bisexual ~ straight	576630	0.391
Asexual ~ queer	10560	<0.0001	Gay ~ queer	21211	<0.0001
Asexual ~ straight	100740	<0.0001	Gay ~ straight	349690	0.296
Bisexual ~ gay	108840	0.835	Queer ~ straight	117050	<0.0001

Table 52 Mann-Whitney U-test for every combination of sexual orientation based on *feminine* ratings

Upon close inspection, it appears that there is only a significant difference in *feminine* ratings when a listener identifies as asexual or queer. Typically, the asexual participants have lower mean ratings than all the other sexuality groups, and fall closer to "Neither agree nor disagree" (or 4) than the other sexual orientation groups. Conversely, the queer participants tend to have the highest mean ratings compared to the other groups and do not rate any sentence lower than "Somewhat agree" (or 5).

However, it is important to note that of the total participant pool, only two participants identified as asexual and four as queer. It is possible that these differences may be due to low sampling numbers. This difference in voting based on sexual orientation should be investigated further, and will be considered in the discussion section below.

Having considered the possible *feminine* rating differences based on sexual orientation, the focus will now turn to ratings for *homosexual*. Initial findings show that while there are some minor differences between the LGBTQ and straight groups in regard to ratings of *homosexual*, there are also some interesting similarities. Figure 25 presents the mean rating for each sentence based on sexual orientation groups.
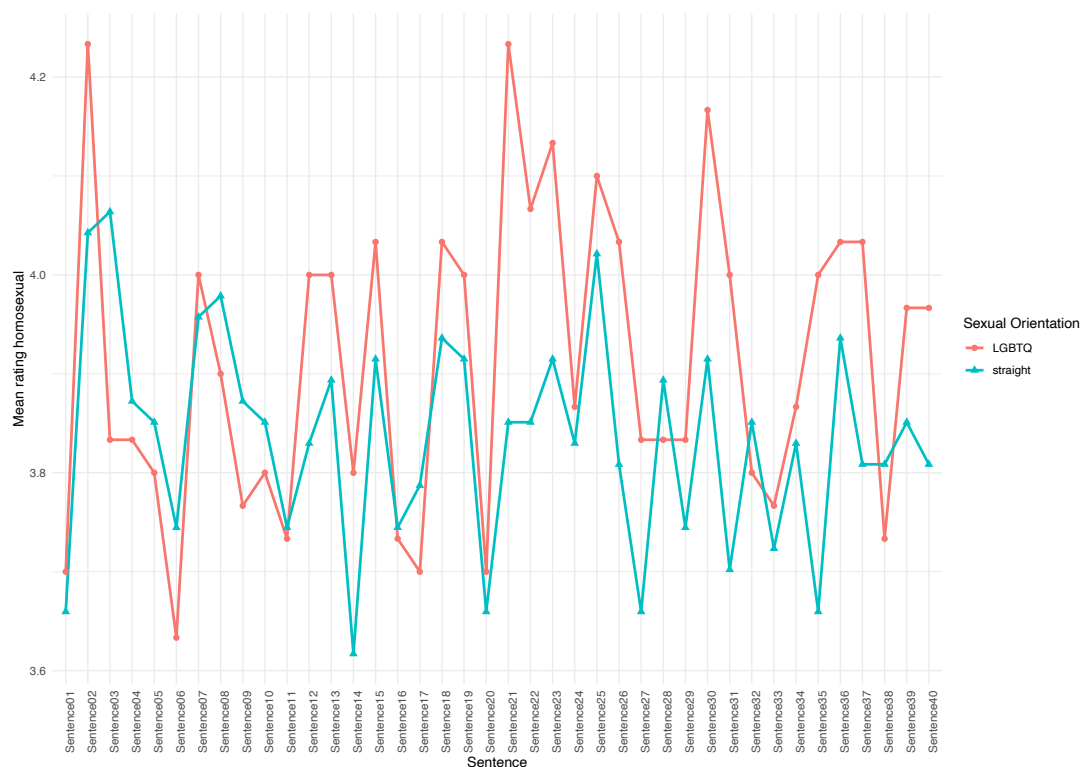
Figure 25 *Homosexual* mean rating grouped by sexual orientation

Table 53 outlines the 10 sentences with the highest ratings for homosexuality for the LGBTQ group and the straight group, as well as how the F0 in that sentence was altered.

LGBTQ (n=30)			Straight (n=47)		
Sentence	**Mean**	**F0 Alteration**	**Sentence**	**Mean**	**F0 Alteration**
2	**4.23**	**-50**	3	4.06	-40
21	4.23	-40	**2**	**4.04**	**-50**
30	**4.17**	**-30**	25	4.02	20
23	**4.13**	**-20**	8	3.98	-20
25	**4.10**	**20**	7	3.96	-40
22	4.07	0	**18**	**3.94**	**-50**
15	**4.03**	**-20**	36	3.94	-50
18	**4.03**	**-50**	15	3.91	-20
26	4.03	-50	**23**	**3.91**	**-20**
36	**4.03**	**-50**	19	3.91	-30
37	4.03	-40	**30**	**3.91**	**-30**

Table 53 Highest *homosexual* ratings by LGBTQ and straight speakers; sentences that appear on both lists emboldened

While there are some differences in the top-rated *homosexual* sentences, and the means are typically higher for the LGBTQ speakers, it is possible to see that both

groups favoured sentences that had had the F0 decreased. There are also 6 sentences out of the 10 total that appear in the highest-rated set for both groups.

The inverse of this pattern can be seen when considering the 10 sentences rated the least *homosexual* by each group; these results are presented in Table 54.

LGBTQ (n=30)			Straight (n=47)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
6	**3.63**	**+50**	14	3.62	+50
1	**3.70**	**0**	**1**	**3.66**	**0**
17	3.70	+20	**20**	**3.66**	**+50**
20	**3.70**	**+50**	27	3.66	+30
11	**3.73**	**+40**	35	3.66	+40
16	**3.73**	**+30**	31	3.70	+30
38	3.73	+30	**33**	**3.72**	**+50**
9	3.77	+30	**6**	**3.74**	**+50**
33	**3.77**	**+50**	**11**	**3.74**	**+40**
5	3.80	+40	**16**	**3.74**	**+30**
10	3.80	+20	29	3.74	+50
14	3.80	+50			
32	3.80	-40			

Table 54 Lowest *homosexual* ratings by LGBTQ and straight speakers; sentences that appear on both lists emboldened

As with the highest-rated sentences, the lowest-rated sentences also have six sentences that appear in each group's set. There is also an inverse pattern with respect to F0 alteration, with the sentences rated least *homosexual* having been digitally altered to increase F0, apart from the first sentence everyone heard. The two groups also appear to have close mean *homosexual* rating ranges, as they did with *feminine*. For *homosexual* ratings, the LGBTQ group has a mean rating range of 0.6, while the straight group has a mean rating range of 0.44.

However, a two-tailed Mann-Whitney U-test suggests that the *homosexual* ratings between the LGBTQ group and the straight group are significantly different ($U = 1168700$, $p = 0.0158$). When investigated further, it was discovered that only two sentences were significantly different based on the LGBTQ and straight groups, while all the other sentences were not significantly different. The results of the Mann-Whitney U-tests for all 40 sentences can be seen in Table 55 below. The two

sentences with significant difference between the groups, sentences 21 and 35, are emboldened.

Sentence	*U*	*p*	Sentence	*U*	*p*
1	714	0.907	**21**	**857.5**	**0.018**
2	816	0.133	22	790.5	0.199
3	614.5	0.222	23	799.5	0.132
4	691.5	0.848	24	707.5	0.976
5	646	0.349	25	721.5	0.804
6	668.5	0.586	26	773.5	0.29
7	733	0.677	27	797	0.177
8	680.5	0.735	28	672	0.623
9	667.5	0.576	29	738.5	0.627
10	670.5	0.586	30	804.5	0.134
11	665.5	0.565	31	822	0.086
12	783.5	0.212	32	689	0.819
13	741.5	0.576	33	704	0.994
14	757.5	0.454	34	756	0.476
15	742.5	0.565	**35**	**833**	**0.041**
16	701	0.957	36	733.5	0.686
17	667	0.58	37	779.5	0.263
18	728.5	0.728	38	681	0.75
19	713.5	0.91	39	718	0.85
20	706.5	0.988	40	744	0.537

Table 55 Mann-Whitney U-test for each sentence based on *homosexual* ratings between LGBTQ and straight participants; emboldened sentences are significantly different

These results suggest that it is exceptional for LGBTQ and straight participant groups to perceive the trait *homosexual* differently from each other based on pitch. Figure 26 charts the two sentences, 21 and 35, that had significantly different ratings between LGBTQ and straight speakers. It is interesting to note that sentence 21 had the second highest rating of *homosexual* for the LGBTQ group but did not appear in the top 10 of the straight group. Conversely, sentence 35 was ranked fifth lowest for *homosexual* by the straight group, but did not appear on the bottom 10 list for the LGBTQ group at all.
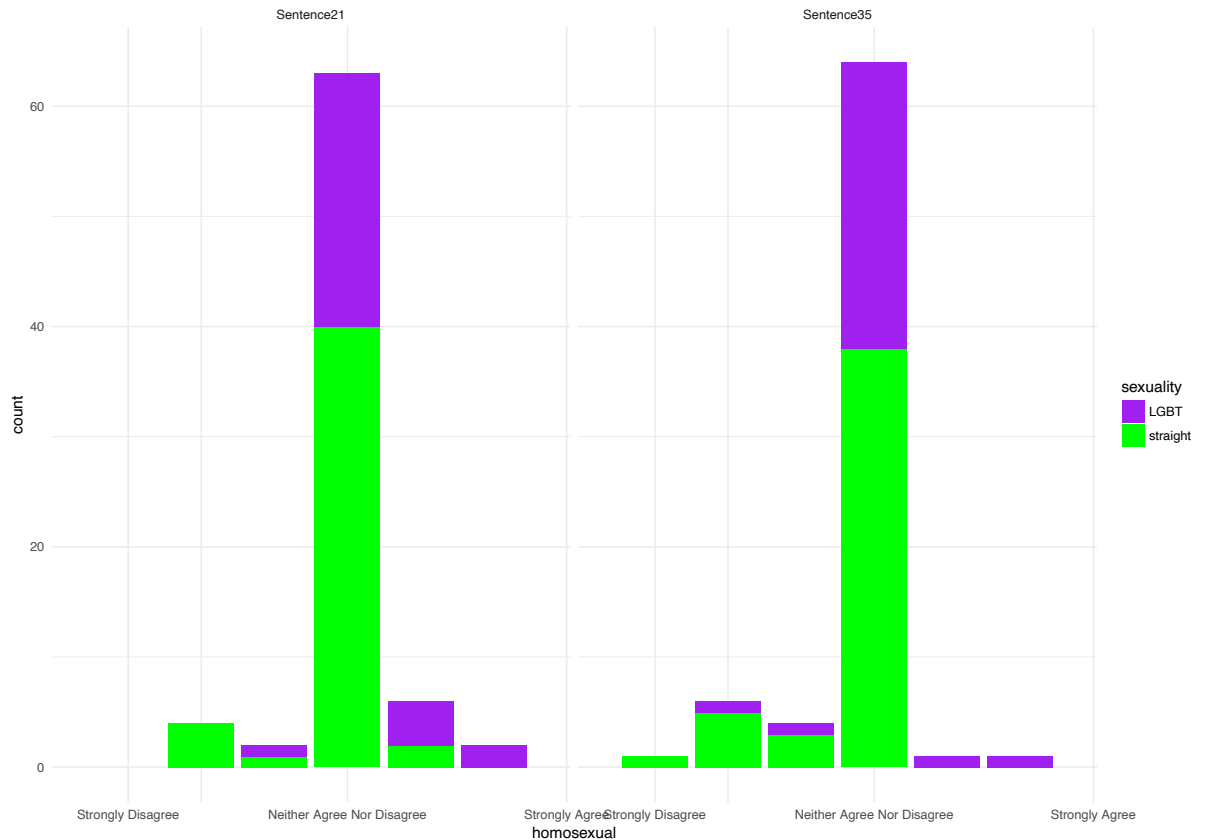
Figure 26 Ratings of *homosexual* for sentences 21 and 35, based on sexual orientation

While it is desirable to consider *homosexual* rating differences based on narrower categories of sexual orientation, as has been done with the trait *feminine*, this analysis is impeded by the number of "neither agree nor disagree" responses for every sentence. With there being only 17 participants that rated at least a quarter of the total sentences something other than "neither agree nor disagree" in regard to *homosexual*, there is not enough data to fully explore how participants with different sexual orientations rated the quality *homosexual* in the present survey.

4.3.2.2 Gender comparisons

Having considered voting differences based on sexual orientation, we now consider rating differences based on listener gender. As with the section on sexual orientation, participants will be grouped together to allow for statistical analysis when there are too few to compare. As only one participant identified as agender and one participant as a trans-man, they will be combined with larger groups. The participant that identified as a trans-man will be included in the male category, and

the participant that identified as agender will be combined with the participants that identified as non-binary under a group of speakers that identify as outside the binary opposition of male and female. This group of agender and non-binary participants will be referred to as "outside the binary" (henceforth OB) so as to not conflate the two. This section will first consider ratings for *feminine*.

A Kruskal-Wallis H-test suggests that there are significant *feminine* rating differences between gender identity groups (Chi square = 67.552, p <0.000039, df = 3). A Mann-Whitney U-test was then run on the three different combinations of gender to test whether all the gender identities rated *femininity* significantly differently from each other. It was found to be different between all three: female and male participants (U = 849920, p < 0.0001), female and OB participants (U = 254070, p < 0.0001), and male and OB participants (U = 113990, p < 0.0001).

While all three genders typically rate the same stimuli sentence in the same direction, they differ by how strongly they rate. Male participants had the highest ratings of *feminine* for all the stimuli apart from one sentence. The OB participants had the lowest *feminine* ratings across all the stimuli, with the exception of two tokens. Finally, female participants rated in between the other two genders. The mean ratings for each gender group can be seen in Figure 27 below.
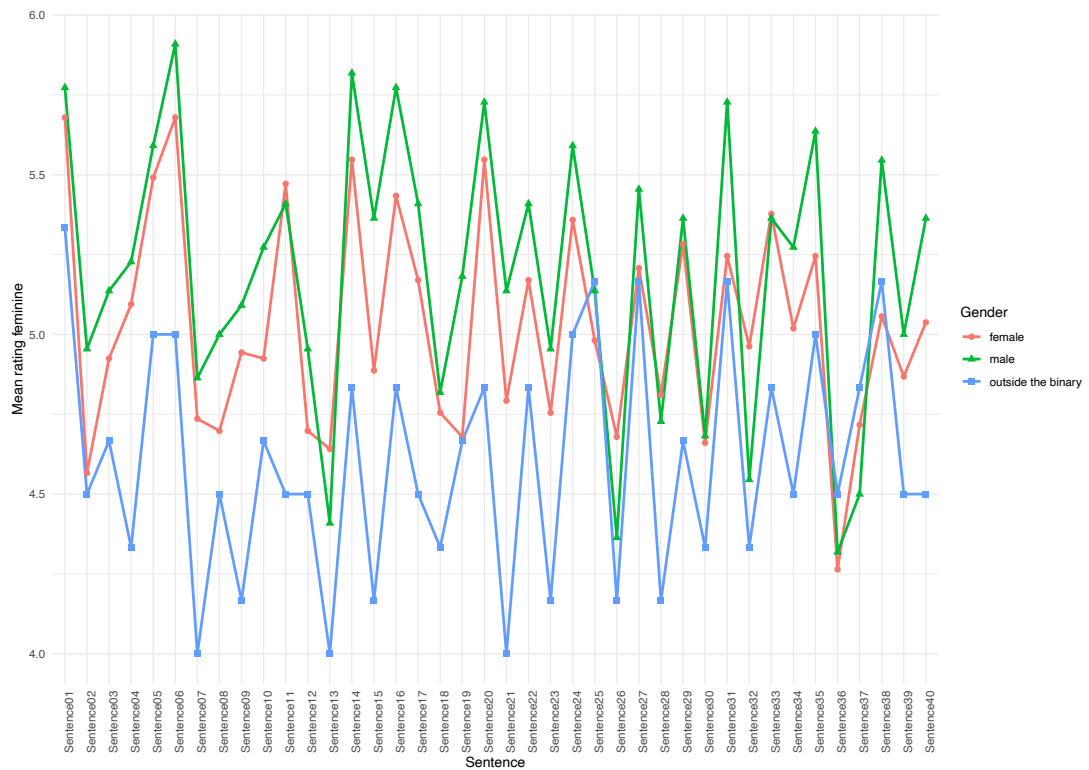
Figure 27 *Feminine* mean rating grouped by gender

Despite these differences, there is still considerable overlap in the sentence ratings. Table 56 below presents the sentences rated highest in the trait *feminine* for all three gender identity categories.

Female (n=53)			Male (n=22)			Outside Binary (n=6)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
1	**5.68**	**0**	6	5.91	+50	**1**	**5.2**	**0**
6	**5.68**	**+50**	14	5.82	+50	25	5.2	+20
14	5.55	+50	**1**	**5.77**	**0**	27	5.2	+30
20	**5.55**	**+50**	**16**	**5.77**	**+30**	31	5.2	+30
5	**5.49**	**+40**	**20**	**5.73**	**+50**	38	5.2	+30
11	5.47	+40	31	5.73	+30	**5**	**5**	**+40**
16	**5.43**	**+30**	35	5.64	+40	**6**	**5**	**+50**
33	5.38	+50	**5**	**5.59**	**+40**	**16**	**5**	**+30**
24	**5.36**	**+40**	**24**	**5.59**	**+40**	**20**	**5**	**+50**
29	5.28	+50	38	5.55	+30	22	5	0
						24	**5**	**+40**
						35	5	+40
						37	5	-40

Table 56 Highest *feminine* ratings by gender identity; sentences that appear on all lists emboldened

This table demonstrates that sentences with an increased F0 are rated highly for *feminine,* as was seen in the overall ratings and the breakdown by sexual orientation. However, there is an interesting difference with the OB group. While the sentences that all three categories agreed on are increased by 40 or 50Hz, the OB group has more top-rated sentences with lower manipulation levels than male and female participants. There is even a sentence that appears in the top-rated sentences that has an F0 decrease. This pattern can also be seen in the lowest rated sentences, seen in Table 57 below.

Female (n=53)			Male (n=22)			Outside Binary (n=6)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
36	4.26	-50	36	4.32	-50	**7**	**4**	**-40**
2	**4.57**	**-50**	**26**	**4.36**	**-50**	**13**	**4**	**-30**
13	**4.64**	**-30**	**13**	**4.41**	**-30**	21	4	-40
30	**4.66**	**-30**	37	4.50	-40	9	4.2	+30
19	4.68	-30	32	4.55	-40	15	4.2	-20
26	**4.68**	**-50**	**30**	**4.68**	**-30**	23	4.2	-20
8	4.70	-20	28	4.73	-20	**26**	**4.2**	**-50**
12	4.70	0	18	4.82	-50	28	4.2	-20
37	4.72	-40	**7**	**4.86**	**-40**	**2**	**4.4**	**-50**
7	**4.74**	**-40**	**2**	**4.95**	**-50**	4	4.4	-30
			12	4.95	0	18	4.4	-50
			23	4.95	-20	**30**	**4.4**	**-30**
						32	4.4	-40
						34	4.4	+20

Table 57 Lowest *feminine* ratings by gender identity; sentences that appear on all lists emboldened

The lowest-rated sentences based on the trait *feminine* also pattern as expected, with sentences with a large F0 decrease being rated as least feminine. However, as with the highest-rated sentences, there is slight variation between female and male participants and the OB group.

While there is a general trend to rate sentences with decreased F0 as less *feminine*, there are more sentences included in the lowest list in Table 57 that have only been decreased by 20 and 30Hz, instead of predominantly being comprised of sentences decreased by 40 and 50Hz. There are also two sentences that break the pattern for the OB group because these sentences have an F0 increase. The differences between the female and male participants and the OB participants can also be seen

in the rating range for *feminine*. The female participants have a mean rating range of 1.42 and the male participants have a mean rating range of 1.59. However, the OB group has a mean rating range of only 0.8, which indicates that this group did not rely on the more extreme responses (i.e. "strongly disagree") as did the other two genders.

The following section will consider how participants rated the quality of *homosexual* based on gender identity categories. It is important to note that there were only two participants in the OB category that rated the sentences as anything other than "neither agree nor disagree" every time. This data will be presented along with the other gender categories, but may not be as revealing due to a small number of participants. Figure 28 below presents the mean rating of each sentence by all genders.
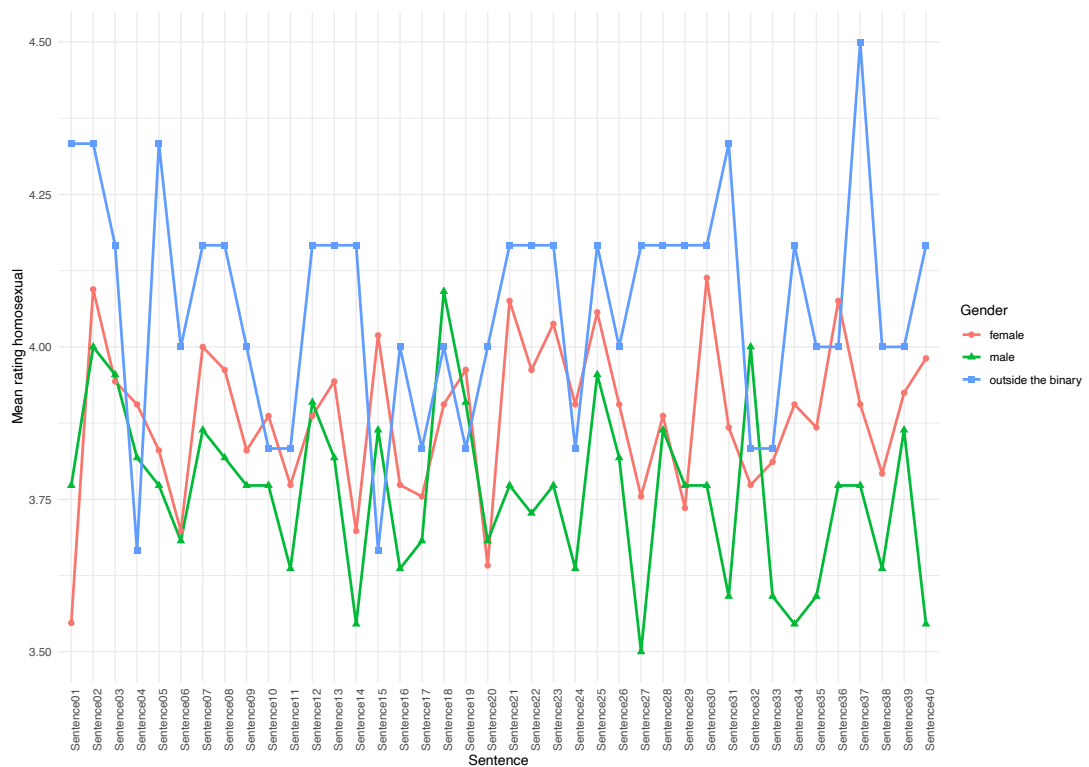
Figure 28 *Homosexual* mean rating grouped by gender

As with the feminine ratings, the male participants are typically more extreme in their rating and they often having the most extreme rating of the gender groups. Also as before, the female participants commonly are more in the middle of the

rating, though as has been mentioned, the OB group did not have many participants and therefore their data is based on a smaller sample. It appears that male participants were more likely to rate using stronger scales than the female participants.

In order to look closer at some of these ratings, Table 58 presents the sentences rated the highest for sounding *homosexual*.

Female (n=53)			Male (n=22)			Outside Binary (n=6)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
30	4.11	-30	18	4.09	-50	37	4.60	-40
2	**4.09**	**-50**	**2**	**4.00**	**-50**	1	4.40	0
21	4.08	-40	32	4.00	-40	**2**	**4.40**	**-50**
36	4.08	-50	3	3.95	-40	5	4.40	+40
25	**4.06**	**+20**	**25**	**3.95**	**+20**	31	4.40	+30
23	4.04	-20	12	3.91	0	3	4.20	-40
15	4.02	-20	19	3.91	-30	**7**	**4.20**	**-40**
7	**4.00**	**-40**	**7**	**3.86**	**-40**	8	4.20	-20
40	3.98	0	15	3.86	-20	12	4.20	0
8	3.96	-20	28	3.86	-20	13	4.20	-30
19	3.96	-30	39	3.86	-30	14	4.20	+50
22	3.96	0				21	4.20	-40
						22	4.20	0
						23	4.20	-20
						25	**4.20**	**+20**
						27	4.20	+30
						28	4.20	-20
						29	4.20	+50
						30	4.20	-30
						34	4.20	20
						40	4.20	0

Table 58 Highest *homosexual* ratings by gender identity; sentences that appear on all lists emboldened

Table 59 presents similar parallels for the sentences rated the lowest for sounding *homosexual*.

Female (n=53)			Male (n=22)			Outside Binary (n=6)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
1	3.55	0	27	3.50	+30	4	3.6	-30
20	3.64	+50	14	3.55	+50	15	3.6	-20
6	3.70	+50	34	3.55	+20	10	3.8	+20
14	3.70	+50	40	3.55	0	**11**	**3.8**	**+40**
29	3.74	+50	31	3.59	+30	17	3.8	+20
17	3.75	+20	33	3.59	+50	19	3.8	-30
27	3.75	+30	35	3.59	+40	24	3.8	+40
11	**3.77**	**+40**	**11**	**3.64**	**+40**	32	3.8	-40
16	**3.77**	**+30**	**16**	**3.64**	**+30**	33	3.8	+50
32	3.77	+40	24	3.64	+40	6	4.0	+50
			38	3.64	+30	9	4.0	+30
						16	**4.0**	**-30**
						18	4.0	-50
						20	4.0	+50
						26	4.0	-50
						35	4.0	+40
						36	4.0	-50
						38	4.0	+30
						39	4.0	-30

Table 59 Lowest *homosexual* ratings by gender identity; sentences that appear on all lists emboldened

The inconsistent ratings of the OB group hint that the participants who voted on the quality of *homosexual* may have chosen their responses randomly. Interestingly, there is a greater *homosexual* mean rating range for the OB participants than there is for the female and male participants. The OB group has a mean *homosexual* range of 1, while female listeners have a mean range of 0.56 and males have a mean range of 0.59.

Further research would have to be conducted in order to establish whether the voting by the OB group was in fact random or whether perceptions work differently among OB participants. However, the female participants in this group appear to be particularly sensitive to large increases of F0 when rating *homosexual*.

4.3.2.3 Regional comparisons

Finally, this section will consider how participants voted according to UK region of origin. Only participants that identified as being from England will be included.

Findings will be presented by categorising participants into three regions: South, Midlands, and North. The South consists of participants from East of England, London, South East, and South West. The Midlands consists of participants from East Midlands and West Midlands. And the North consists of participants from Yorkshire and the Humber, North East, and North West.

A Kruskal-Wallis H-test suggests that there are significant *feminine* rating differences between English regions (Chi square = 199.21, p <0.00001, df = 2). A Mann-Whitney U-test was then run on the three different region groups to discover whether they rated *femininity* significantly differently from each other: South and Midlands (U = 142400, p = 0.09713), South and North (U = 420160, p < 0.0001), and Midlands and North (U = 199820, p < 0.0001). These results suggest that the participants from the North have significantly different ratings for *feminine* than from either the South or the Midlands, which are not significantly different from each other.

All three regions demonstrated similar patterns, with stimuli with increased F0 alterations typically being rated higher on the *feminine* scale and stimuli with decreased F0 alterations being rated lower on the *feminine* scale. However, the North differs to the South and Midlands due to how high *feminine* was rated consistently, across all the stimuli. Participants from the North were more likely to rate all the stimuli as more feminine than the other regions. This can be seen in Figure 29 below.
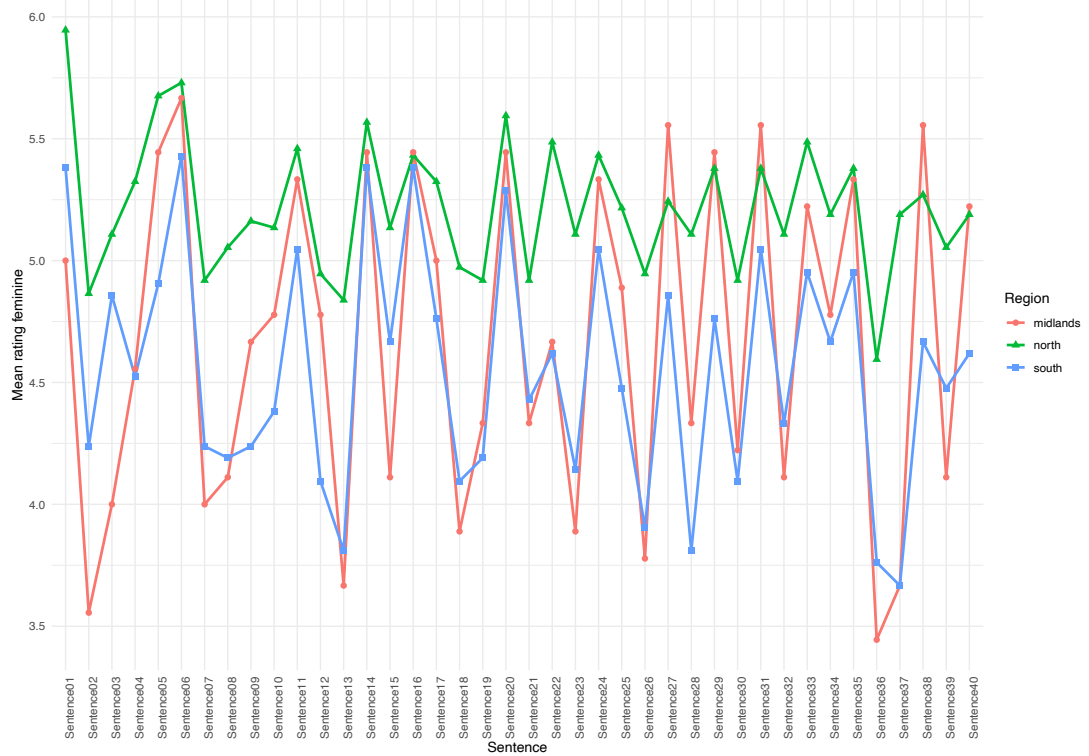
Figure 29 *Feminine* mean rating grouped by region

To consider these mean ratings more clearly, Table 60 presents the highest rated *feminine* sentences for the region groups.

South (n=21)			Midlands (n=9)			North (n=37)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
6	**5.43**	**+50**	**6**	**5.67**	**+50**	1	5.95	0
1	5.38	0	27	5.56	+30	**6**	**5.73**	**+50**
14	**5.38**	**+50**	31	5.56	+30	5	5.68	+40
16	**5.38**	**+30**	38	5.56	+30	**20**	**5.59**	**+50**
20	**5.29**	**+50**	5	5.44	+40	**14**	**5.57**	**+50**
11	**5.05**	**+40**	**14**	**5.44**	**+50**	22	5.49	0
24	**5.05**	**+40**	**16**	**5.44**	**+30**	33	5.49	+50
31	5.05	+30	**20**	**5.44**	**+50**	**11**	**5.46**	**+40**
33	4.95	+50	29	5.44	+50	**16**	**5.43**	**+30**
35	4.95	+40	**11**	**5.33**	**+40**	**24**	**5.43**	**+40**
			24	**5.33**	**+40**			
			35	5.33	+40			

Table 60 Highest *feminine* ratings by region groups; sentences that appear on all lists emboldened

Inverse ratings can be seen in Table 61, below, which presents results for the lowest rated *feminine* sentences for the region groups.

South (n=21)			Midlands (n=9)			North (n=37)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
37	3.67	-40	**36**	**3.44**	**-50**	**36**	**4.59**	**-50**
36	**3.76**	**-50**	2	3.56	-50	**13**	**4.84**	**-30**
13	**3.81**	**-20**	**13**	**3.67**	**-30**	2	4.86	-50
28	3.81	-20	37	3.67	-40	7	4.92	-40
26	**3.90**	**-50**	**26**	**3.78**	**-50**	19	4.92	-30
12	4.10	0	**18**	**3.89**	**-50**	21	4.92	-40
18	**4.10**	**-50**	23	3.89	-20	30	4.92	-30
30	4.10	-30	3	4.00	-40	12	4.95	0
23	4.14	-20	7	4.00	-40	**26**	**4.95**	**-50**
8	4.19	-20	8	4.11	-20	**18**	**4.97**	**-50**
19	4.19	-30	15	4.11	-20			
			32	4.11	-40			
			39	4.11	-30			

Table 61 Lowest *feminine* ratings by region groups; sentences that appear on all lists emboldened

As can be seen in Tables 60 and 61, there are many similarities in the ratings between the regions and they follow similar patterns, with higher-rated *feminine* sentences having increased F0 and lower-rated *feminine* sentences having decreased F0. Yet each group has a very different range from one another. The mean *feminine* rating range for the northern participants is 1.36 and the mean *feminine* rating range for the southern listeners is 1.76. The participants from the midlands have a high mean *feminine* rating range of 2.23.

However, what is compelling between the groups is how strongly participants voted. Participants from both the South and Midlands had much lower mean ratings than the northern group, a trend which can be seen particularly in Table 61 and the lower ratings. The South and Midlands listeners both have lowest ratings under 3.7, while listeners from the North never rate a sentence under 4.5. There appear to be significant differences in perceptions of femininity based on English regions.

Results will now be presented based on how participants from different regions rated the quality of *homosexual* for the survey. A Kruskal-Wallis H-test suggests that there are significant *homosexual* rating differences between English regions (Chi square = 11.245, $p = 0.003616$, df = 2). A Mann-Whitney U-test was then run on the three different regions to discover if the regions rated *homosexual* significantly different from each other. The results show the following: South and

Midlands ($U = 149020$, $p = 0.6049$), South and North ($U = 652540$, $p = 0.004951$), and Midlands and North ($U = 282970$, $p = 0.008573$). As before, these results suggest that the participants from the North have significantly different ratings for *homosexual* from those given by listeners from the South and the Midlands, which are not significantly different from each other.

As before with ratings of *feminine*, participants from the North rate *homosexual* differently from those in the South and Midlands. While the South and Midlands have mean ratings that vary more by stimuli, participants in the North average a narrower difference across all of the stimuli sentences and generally do not rate *homosexual* very high. This can be seen in Figure 30 below.
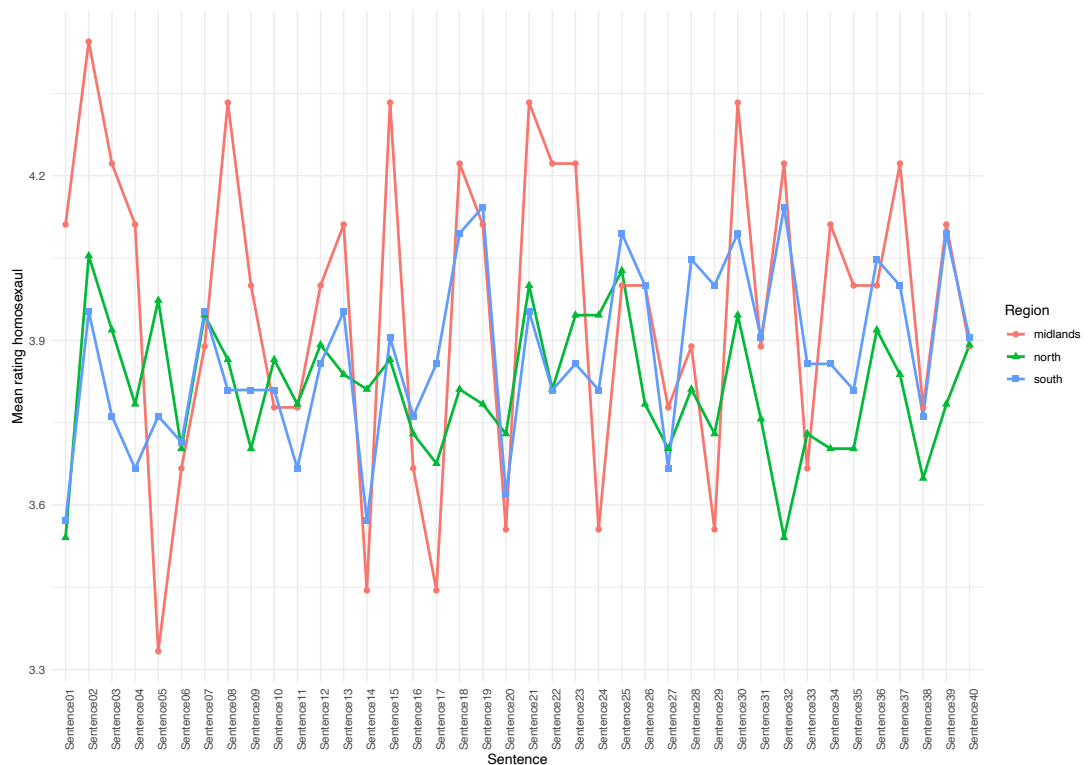
Figure 30 *Homosexual* mean rating grouped by region

Table 62 presents the sentences rated the highest for *homosexual* based on each region to allow for a closer examination of these differences.

South (n=21)			Midlands (n=9)			North (n=37)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
19	4.14	-30	2	4.44	-50	2	4.05	-50
32	4.14	-40	8	4.33	-20	25	4.03	+20
18	4.10	-50	15	4.33	-20	21	4.00	-40
25	4.10	+20	21	4.33	-40	5	3.97	+40
30	**4.10**	**-30**	**30**	**4.33**	**-30**	7	3.95	-40
39	4.10	-30	3	4.22	-40	23	3.95	-20
28	4.05	-20	18	4.22	-50	24	3.95	+40
36	4.05	-50	22	4.22	0	**30**	**3.95**	**-30**
26	4.00	-50	23	4.22	-20	3	3.92	-40
29	4.00	+50	32	4.22	-40	36	3.92	-50
37	4.00	-40	37	4.22	-40			

Table 62 Highest *homosexual* ratings by region groups; sentences that appear on all lists emboldened

While so far we have generally seen that sentences which are rated higher for the quality of *homosexual* have decreased F0, there are some sentences included in Table 62 that have an increased F0. This may be due to the fact that the ratings generally are quite low as a whole. For the northern group, only three sentences have a rating of four or above, which is only "neither agree nor disagree". This means that participants from the North frequently tended to disagree that the voice sounded *homosexual* regardless of the degree of F0 alteration. Table 63 shows that some of the more inconsistent ratings are also given to the lowest-rated sentences.

South (n=21)			Midlands (n=9)			North (n=37)		
Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration	Sentence	Mean	F0 Alteration
1	3.57	0	5	3.33	+40	1	3.54	0
14	3.57	+50	14	3.44	+50	32	3.54	-40
20	**3.62**	**+50**	17	3.44	+20	**38**	**3.65**	**+30**
4	3.67	-30	**20**	**3.56**	**+50**	17	3.68	+20
11	3.67	+40	24	3.56	+40	**6**	**3.70**	**+50**
27	**3.67**	**+30**	29	3.56	+50	9	3.70	+30
6	**3.71**	**+50**	**6**	**3.67**	**+50**	27	3.70	+30
3	3.76	-40	**16**	**3.67**	**+30**	34	3.70	+20
5	3.76	+40	33	3.67	+50	35	3.70	+40
16	**3.76**	**+30**	10	3.78	+20	**16**	**3.73**	**+30**
38	**3.76**	**+30**	11	3.78	+40	**20**	**3.73**	**+50**
			27	**3.78**	**+30**	29	3.73	+50
			38	**3.78**	**+30**	33	3.73	+50

Table 63 Lowest *homosexual* ratings by region groups, sentences that appear on all lists emboldened

To judge by the results shown in Tables 62 and 63 it appears that English participants rated the degree to which the speaker sounded *homosexual* consistently low throughout the survey. However, there is more agreement between the three groups on sentences that appear in the lowest ratings than there is for the highest ratings. Yet the range of ratings used by the groups is again slightly different. The northern speakers have a mean *homosexual* rating range of 0.51 and the southern speakers a range of 0.57. As with ratings of *feminine*, the Midlands speakers have the largest *homosexual* rating range of 1.11.

4.3.2 Quality Correlations

Having considered how different groups of participants rated the key qualities of femininity and homosexuality, it is important to consider how these qualities, and others, correlate throughout the survey.

Table 64 presents the Spearman correlation results for all of the possible combinations of the seven qualities surveyed. Following the table, there will be a more in-depth consideration for some of the key qualities and results of correlation tests run on the data for different groups of speakers.

Quality correlations	r	p
Feminine and friendly	0.3157	<0.0001
Feminine and homosexual	-0.1409	<0.0001
Feminine and intelligent	0.2395	<0.0001
Feminine and low pitch	-0.3461	<0.0001
Feminine and speaker age	-0.1811	<0.0001
Feminine and trustworthy	0.2453	<0.0001
Friendly and homosexual	-0.0626	0.0004
Friendly and intelligent	0.3243	<0.0001
Friendly and low pitch	-0.2232	<0.0001
Friendly and speaker age	-0.1516	<0.0001
Friendly and trustworthy	0.56952	<0.0001
Homosexual and intelligent	**0.0009**	**0.9574**
Homosexual and low pitch	0.1429	<0.0001
Homosexual and speaker age	0.07317	<0.0001
Homosexual and trustworthy	-0.0539	0.0022
Intelligent and low pitch	**-0.0052**	**0.7662**
Intelligent and speaker age	0.0837	<0.0001
Intelligent and trustworthy	0.4291	<0.0001
Low pitch and speaker age	0.4196	<0.0001
Low pitch and trustworthy	-0.116	<0.0001
Speaker age and trustworthy	-0.0438	0.0127

Table 64 Results of Spearman correlation tests; emboldened sections show no significant correlation

From Table 64, it is clear to see that there are many correlations between the qualities participants were asked to rate. The only two combinations of qualities that are not statistically significantly correlated are *homosexual* and *intelligent* and *low pitch* and *intelligent*. As intelligence was one of the filler traits, it is not relevant to the current study to consider why this is the only trait that did not have significant correlations to every other quality; this will not be considered further.

There are some correlations that appear to be less arbitrary than others: for example, there is a strong positive correlation between perception of *low pitch* and *speaker age* ($r = 0.4196$, $p = < 0.0001$). Sentences that are rated more highly for *low pitch* are also perceived as having been produced by an older speaker and conversely the less a sentence is rated for *low pitch*, the younger the speaker is assumed to be. Similarly, as "feminine speech" is often associated with higher pitch (Munson, 2007a; Fraccaro *et al.*, 2011), it is logical that there is a significant negative correlation between femininity and *low pitch* perceptions ($r = -0.3461$, $p < 0.0001$).

However, there are some unexpected correlations that appear in this data. Femininity is positively correlated with *friendliness* (r = 0.3157, p < 0.0001), intelligence (r = 0.2395, p < 0.0001), and *trustworthiness* (r = 0.2453, *p* < 0.0001). These findings would benefit from further investigation to understand in which direction, if any, these qualities may be influencing each other. For example, is a speaker perceived as being friendlier because she is perceived as more feminine? Is she perceived as more feminine because she seems to be more friendly? Or are these perceptions completely unrelated? The discussion below will consider future research based on these findings.

While the results show a negative correlation between femininity and homosexuality when considering the entire data set, there is a weaker correlation when considering individual sentences. Of the 40 sentences in the survey, only five showed a significant negative correlation between femininity and homosexuality. Table 65 presents the results of Spearman's correlation tests for femininity and homosexuality for every sentence.

Sentence	F0 Alteration	r	p	Sentence	F0 Alteration	r	p
1	0	-0.0977	0.3857	21	-40	-0.0856	0.4475
2	**-50**	**-0.3875**	**0.0004**	22	0	-0.1193	0.2886
3	-40	-0.1702	0.1288	23	-20	-0.0252	0.8235
4	-30	-0.0058	0.9592	**24**	**40**	**-0.2557**	**0.0212**
5	40	0.0007	0.9947	25	20	-0.0593	0.5990
6	50	-0.0153	0.8922	26	-50	-0.1528	0.1732
7	**-40**	**-0.2566**	**0.0208**	27	30	-0.1012	0.3686
8	-20	-0.1167	0.2996	28	-20	-0.0469	0.6777
9	30	-0.0358	0.7508	29	50	-0.1904	0.0886
10	20	-0.0153	0.8924	30	-30	-0.1824	0.1032
11	40	0.0245	0.8279	31	30	-0.1180	0.2943
12	0	-0.2010	0.0720	32	-40	-0.0965	0.3913
13	-30	-0.0771	0.4941	33	50	0.0964	0.3921
14	50	-0.1062	0.3452	34	20	-0.0877	0.4363
15	**-20**	**-0.2545**	**0.0218**	35	40	-0.1653	0.1402
16	**30**	**-0.2273**	**0.0412**	36	-50	-0.1154	0.3051
17	20	-0.0443	0.6945	37	-40	-0.1404	0.2113
18	-50	-0.0517	0.6465	38	30	-0.1714	0.1260
19	-30	-0.2000	0.0734	39	-30	-0.2048	0.0667
20	50	-0.0422	0.7083	40	0	-0.1128	0.3162

Table 65 Results of Spearman correlation tests for ratings of *femininity* and *homosexuality*; emboldened figures show significant correlation

There are no distinct patterns with respect to sentences for which significant femininity/homosexuality correlations were found and those for which there appeared to be no such relationship. Interestingly, only three sentences show any positive correlation (sentences 5, 11, and 33) and the rest are all negatively correlated. This may account for the fact that there is a significant negative correlation between femininity and homosexuality when considering all of the data.

4.4 Discussion

4.4.1 Participant interaction with the survey

The results section began by presenting just how frequently participants in the survey voted "neither agree nor disagree" when rating the quality of *homosexual*. In fact, this rating was so common that half the participants used this rating for all 40 sentences when asked about homosexuality. The frequency of this rating indicates that this was not random chance, but something done purposefully by a large portion of the participants. There are two possibilities for why participants voted in this way.

The first possibility is that participants genuinely did not have any preconceived notions of a female gay voice and were unsure how to rate any given sentence. As there are apparently not many stereotypes for participants to rely on when judging whether a voice fulfils the "gay voice" criterion, some participants may not have any strong idea of what makes one voice sound more or less gay than another. Munson (2007a) also comments on the fact that there are minimal popular-culture references to gay voice for women. If it is the case that participants do not have a frame of reference, then their neutrality in answering may come from not having any answer to give, as there are no stereotypes to rely on.

However, a second reason participants may have voted neutrally throughout the survey was to do with a fear of being culturally insensitive. Generally, it is viewed negatively if one relies upon stereotypes to identify or judge a person one does not know, and this survey asked participants directly to make these kinds of judgements. Nelson, Acker, and Manis (1996, p. 15) write in their study on stereotypes: "Stereotypes are typically reviled because their content is often negative and

154

inaccurate. But a further hazard of stereotypes is that they often beget *overgeneralizations*, or the blanket application of a trait to virtually all members of a group" (emphasis in original work). For this reason, even if the stereotype does not have inherently negative connotations, it can still be harmful based on the fact that it is an overgeneralisation.

Nelson *et al.* also write: "Those who would hope to minimize the deleterious social consequences of stereotyping suggest that we must control our impulse to judge others on the basis of broad social category membership, and should instead focus on the unique attributes of the individual" (ibid. p14). In the present study, by asking participants to gauge whether the speaker had a gay voice or not, they may have felt they would be forced to rely on stereotypes that participants were uncomfortable about being seen to endorse. The best way to avoid this situation was not to pass any judgement at all.

Livia (2002) also addressed the difficulty participants face when asked to comment on cultural stereotypes. Livia recalls asking participants about their image of a *camionneuse*, a derogatory French word equivalent to the phrase in English "diesel dyke". She writes that by asking participants to describe a person that would be a *camionneuse*, the participants were in "a catch 22: if they were able to elaborate the cultural connotations of the term *camionneuse* then they reinforced irksome gender stereotypes. Yet if they failed at the task, they appeared to be cultural illiterates" (Livia, 2002, p. 96)

A similar catch 22 is also present in the current research. Participants may work with a stereotype of a voice they associate with gay women, but to acknowledge this would mean being complicit with the stereotype. As was seen in the production study in Section 3.3.1.3, those participants also acknowledged that there may be a stereotypical gay voice for women, but were hesitant to make any strong claims about it. By giving a neutral answer in this survey, participants did not have to address cultural stereotypes of a gay voice and could comfortably navigate away from the potentially culturally insensitive situation.

Yet this reason for potential resistance to addressing the quality of *homosexual* is further complicated by the fact that participants were regularly willing to indicate whether they thought the speaker sounded *friendly*, *intelligent*, or *trustworthy*. Friendliness, intelligence, and trustworthiness are all judgement-based traits and the participant was assessing how much the speaker has these qualities, which would have implied social value. Rating a speaker as more or less *friendly*, *intelligent*, or *trustworthy* could be received by the speaker as a compliment or insult depending on the rating. As with *homosexual*, there are wider social implications when a speaker is rated as not being friendly, for example, or when rating a speaker as being exceptionally intelligent.

While there were a few participants that avoided these social judgments and voted "neither agree nor disagree" for *intelligent* and *trustworthy*, the number of people that voted in this way was drastically lower than those that solely rated "neither agree nor disagree" for the quality of *homosexual*. This indicates that participants generally feel more comfortable rating a person for perceived intelligence, but not perceived homosexuality. Further research should consider why participants would view a judgement of friendliness, intelligence, or trustworthiness to be more socially acceptable than judging a person's sexual orientation.

I argue that it is the combination of both of the above reasons that explains why participants were unlikely to give the sentences non-neutral ratings for *homosexuality*. In the absence of a well-known stereotype to base the rating on, participants may be unlikely to give a strong rating, particularly if it is seen as socially unacceptable to do so, and especially when explicitly asked. Further research should continue to explore this topic and investigate the possibility whether there are certain groups of listeners that are more likely to make female gay voice judgements than others. Similarly, continued research on the male gay voice (which includes work by Smyth, Jacobs and Rogers, 2003; Levon, 2006; Sulpizio *et al.*, 2015) may indicate that participants are more likely to make ratings, even if it makes them feel socially uncomfortable, if there are more distinct and well-known stereotypes.

Having considered whether participants were comfortable rating sentences in a general sense, the focus will now be placed on how the sentences were rated and what these ratings could imply about the perception of pitch in the speech of a female speaker, particularly in association with perceptions of femininity and homosexuality.

4.4.2 F0 and perceived qualities

One of the immediately noticeable patterns in the responses given by the participants in this survey was that they appeared to be sensitive to F0. When rating how low-pitched the speaker's voice sounded, participants consistently rated sentences with a decreased F0 as lower in pitch and sentences with an increased F0 as less low-pitched. Participants were even sensitive to the degree of F0 alteration. The four lowest-rated sentences for low pitch were all increased by 50Hz, which was the maximum any sentence stimuli was altered by. Generally, the mean rating for low pitch coincided with the extent to which the F0 was altered for that sentence. The most common time the ordering begins to falter is when the alteration is increased or decreased by 20Hz or the sentence was unchanged. This data suggests that participants are sensitive to pitch changes when they are specifically asked to consider them.

These results appear to be in stark contrast to the results of Tompkinson and Watt (2018), and are worth considering further. In their research, Tompkinson and Watt found only relatively weak correlations between participants' perception of pitch and average F0. However, in the Tompkinson and Watt study, the stimuli were created using speech produced by 12 different speakers, and included three different accents of English. This indeed is a key difference between their study and the present perception study. As Honorof and Whalen (2005) suggest, "the perception of linguistic pitch targets may become possible as a function of exposure to a speaker's voice" (p21930). In the present study, the participants would have had an opportunity to become accustomed to the speaker, even if they did not realise it was the same person. The responses to low pitch in the present study, by comparison with the Tompkinson and Watt study, may indicate that participants are more accurate with pitch perception after prolonged exposure to a speaker.

A second explanation for the accuracy of the pitch perception in the present study may be that every participant initially heard a sentence with a mean pitch of ~200Hz, which acted as the reference point for all the stimuli they heard afterwards. By having this initial sentence to compare the following sentences against, participants may have been given an advantage and were then more accurate in their perception.

This same sensitivity also appeared when participants were asked to rate the speaker's age. Sentences that had an increased F0 were consistently rated as being spoken by a younger speaker than was the case for sentences with decreased F0. Participants were clearly relying on F0 as an indication of age when rating the speaker. While there was more variability in age ratings and the amount of F0 alteration when perceiving older speakers, there was still a clear pattern that sentences with a decreased F0 were deemed to have been produced by older speakers than sentences with an increased F0.

On the basis of these two qualities it is possible to say that participants are sensitive to F0 and can even be aware of F0 changes between samples of as little as 10Hz. This clear awareness of pitch allows the results to be more confidently interpreted, as it indicated that participants were not rating randomly.

In the current data, there is only one quality that is not significantly correlated with F0: *trustworthy*. This lack of correlation may relate to internally conflicting findings of O'Connor and Barclay (2017), who found that higher-pitched female voices were perceived as more trustworthy in an economic context, but lower-pitched voices were more *trustworthy* in general. It is possible that there may be other voice qualities that may be associated with levels of trustworthiness for women's speech and F0 is not as significant in this judgement. However, as this quality was purely used as a filler in the survey, it is beyond the scope of this thesis to consider further vocal qualities related to how trustworthy a speaker is perceived to be.

The two most important traits in the present study are *feminine* and *homosexual*. Ratings of femininity were significantly positively correlated with F0 alteration,

158

whereby sentences with higher F0 values consistently being rated as more *feminine* than sentences with lower F0 values. This is in line with the findings of previous work (Munson, 2007a; Feinberg *et al.*, 2008; Fraccaro *et al.*, 2011) which found that F0 is positively correlated with femininity. The correlations between high F0, femininity, and other traits indicate potentially large-scale social ramifications. This will be discussed in more detail in the trait correlations section under perceived femininity, (4.4.4.1).

Finally, despite the frequency of ratings of "neither agree nor disagree" for the trait of *homosexual*, participants still displayed a general pattern of associating decreased F0 with higher levels of *homosexuality*. The two factors emerged as being significantly negatively correlated. When speakers do report perceiving any sort of gay voice, it is typically associated with a lower F0.

This is in contrast to the research by Sulpizio *et al.* (2019), who did not find a consistent significant correlation between female speaker's mean F0 and perceived sexual orientation, and only found the correlation for the combination of Portuguese speakers with German listeners. However, these differences in findings may be explained in two ways. First, the research conducted by Sulpizio *et al.* used stimuli that were spoken by Italian, Portuguese, and German speakers and the participants in this study were also Italian, Portuguese, and German. It is possible that F0 may be a more significant signal of a female gay voice in English than it is in these other languages.

A second reason for the present results differing from Sulpizio *et al.'s* research may relate back to Tompkinson and Watt's (2018) research discussed above. In both of the latter studies, listeners were asked to respond to the voices of multiple speakers. However, while listeners in the present study were led to believe they would be hearing the voices of multiple speakers, in fact there was only one. It may be that when all other vocal qualities are controlled for, participants then rely on pitch as an indication of a female gay voice. Yet, if other diverse vocal qualities are present, these qualities influence the perception of a gay voice and pitch is no longer a reliable factor.

However, as the focus of the present study is to understand whether F0 has any influence on perceptions of a gay voice for women, the methodology used is ideal for investigating that one feature in a targeted manner. Future studies that focus solely on other targeted features may reveal the bundle of linguistic factors that, in combination, constitute a gay voice for female speakers.

4.4.3 Group comparisons on femininity and homosexuality

While there were interesting patterns based on general trends, it is also important to understand how certain groups rated the data, as the different identities of the participants may have had a strong impact on how they interpreted the stimuli. This consideration of certain identity factors about the participants has not commonly been considered in previous work on a gay voice for either male or female speakers.

Previous perceptual work on the production of /s/ among male speakers by Mack and Munson (2012) acknowledges that participants were not asked about their sexual orientation and the researchers assumed that participants "represent an average cohort of middle-SES traditional college-aged adults" (p202). In his study of perceptions of a male gay voice, Levon (2006) does not document details of the participants, apart from gender. Smyth, Jacobs, and Rogers (2003) do have one group of gay male participants and then a second group of mixed participants, but these mixed participants were only asked about their gender and not their sexual orientation. In the research conducted by Sulpizio *et al.* (2015) and Sulpizio *et al.* (2019), participants that did not identify as straight were excluded as there were not enough participants to allow for statistical comparison. However, these studies did consider in-group identity based on participants sharing the same language as the stimuli they were listening to.

The present work considers not just gender differences, but also sexual orientation and regional differences. Though some of the identity groups are quite small in the present study, these groups still provide interesting insights into how listener identity may influence their perceptions of a gay voice for women.

4.4.3.1 Sexual orientation comparisons

The results indicate a lack of significant differences in perceptions of femininity across larger sexual orientation listening groups. The general groupings of LGBTQ and straight participants did not differ significantly when rating levels of *femininity* for the speaker. However, when the listener sample was broken down according to narrower definitions of sexual orientation, it was found that asexual and queer participants did rate femininity significantly differently from any other sexual orientation group. As acknowledged in the results section, there were few participants that identified as either asexual or queer (2 asexual participants and 4 queer participants), which might have led to the differences in rating. Further research that considers the sexual orientation of participants may reveal whether there is in fact a difference in perceived femininity for these groups of listeners, or if the apparent difference was due to low participant numbers.

As the concept of femininity was not defined for participants before they performed the rating task, and they were expected to follow their own intuition with respect to femininity, it may be possible that participants that identified as asexual or queer do have different concepts of femininity and would rate the sentences differently from the other participants. Future research may consider how individual participants conceive of femininity.

When comparing homosexuality ratings, it is first evident that the LGBTQ participants tend to favour higher ratings for *homosexual* than straight participants. While both groups have lowest mean ratings that are similar, the LGBTQ participants have higher mean ratings for their top sentences than the straight group. It is possible that the LGBTQ group was more comfortable rating a voice as sounding *homosexual* than the straight group, and would rate it more clearly as such than participants. This range in rating could be due to in-group associations that the straight participants do not have.

The lower ratings given by straight participants may also connected to the "straight categorization bias" (Lick and Johnson, 2016), discussed in the methodology section. Straight participants may have a stronger sense of this straight

categorization bias and are less likely to give high ratings of homosexuality. However, as LGBTQ participants may be more inclined to consider other sexualities than straight ones, they may be more comfortable rating homosexuality higher.

Despite LGBTQ participants rating levels of homosexuality higher than straight participants, the two groups did not differ greatly in the sentences they rated as more or less *homosexual*. While the Mann-Whitney U-test initially indicated that LGBTQ and straight speakers rated significantly differently from each other, with further investigation it was discovered that there were only two sentences that were rated significantly differently. This research suggests that there is not in fact a difference in homosexuality perceptions between LGBTQ and straight people based on speaker pitch.

However, as discussed in the results section, comparisons in ratings across more narrowed definitions of sexual orientation was not possible. With a larger pool of participants, there may be patterns that emerge based on the sexual orientation of the listener and how they perceive homosexuality in a speaker.

4.4.3.2 Gender comparisons

While the effect of participant sexual orientation is not consistently strong in the perceptions of this survey, participant gender does appear to have a bigger effect. For ratings of femininity, Mann-Whitney U-tests found that all three gender categories rated the speaker significantly differently from each other.

Though there were significant differences between female and male participants, the participants outside the gender binary (OB) had the largest rating differences of any group. When rating femininity, OB participants had the smallest mean range at 0.8 between the highest-rated sentence and the lowest-rated sentence. This group also did not appear to be as sensitive to F0 alterations as the male and female participants.

For many groups, the highest- and lowest-rated sentences had F0 alterations of either 40Hz or 50Hz, either increased or decreased depending on the trait (i.e.

increased F0 alteration for *feminine* and decreased F0 alteration for *low pitch*). However, when rating femininity, the OB group did not favour these larger alterations strongly. Of the 10 sentences rated most highly for *feminine* by female participants, eight were sentences with the F0 increased by 40 or 50Hz and male participants had seven sentences increased by 40 or 50Hz. Interestingly, only 5 out of the top 13 highest-rated *feminine* sentences from the OB group had an increased F0 of 40 or 50Hz.

The same pattern appears with the lowest-rated sentences. Female participants had five sentences out of their ten lowest-rated sentences with F0 decreases of 40 or 50Hz and male participants had seven out of 12. OB participants have only five out of fourteen sentences in their lowest mean rating that had F0 decrease of 40 or 50Hz.

Based on both the mean voting range and the sentences that make up the highest- and lowest-rated, there appear to be significant differences for OB participants. These differences also appear when considering the ratings for *homosexual*.

When rating *homosexual*, the OB participants have the largest range in voting (1 between the highest and lowest rated sentences), while female participants have a range of only 0.56 and males of 0.59. Yet the more drastic differences are seen in the sentences that are rated as the most and least *homosexual*.

Male and female participants tend to follow the typical pattern seen throughout the data, with the lowest-rated *homosexual* sentences having an increased F0 and the highest-rated *homosexual* sentences having a decreased F0, and the OB group having much more inconsistent ratings. Though OB participants still show signs of following the pattern that the male and female participants follow, there are more exceptions in their ratings.

As hypothesised above that asexual and queer participants may have different assumptions about femininity, from the results it is possible that OB participants also have different perceptions of femininity as well as homosexuality. The similarities between asexual and queer participants and OB participants are not

entirely random, as none of the OB participants identified as straight. Of the OB group, one participant identified as queer, one as asexual, one as gay, and three as either bisexual or pansexual.

More data would need to be collected from participants who are outside the gender binary and who identify with different sexual orientations, but the present research suggests that there are different perceptions of both femininity and homosexuality among these groups of people.

4.4.3.3 Regional comparisons

The last group comparisons this study investigated were the influence English regions may have on participant perceptions. As acknowledged in Section 2.4, this comparison was made because of the strong social stereotypes there are of northern and southern England and strong perceptions of Yorkshire in particular (Fletcher, 2012). According to Hiraga (2005), rural West Yorkshire accents are generally perceived as sociable, sincere, friendly, comforting and reliable, but have lower ratings for education, intelligence, wealth, successfulness, and elegance. Northern English generally has been found to be perceived to have lower status, according to Giles (1970). This study was interested in discovering if these stereotypes had an effect on the present research.

Participant region had a strong effect on the perceptions of femininity in the data. As with many groups already discussed, the general voting trend was matched for the southern, midland, and northern speakers in this data: sentences rated as more *feminine* had higher F0 alterations and sentences rated as less *feminine* had F0 decreases. However, it is the degree to which the sentences were rated that show the drastic differences between the regions.

The South and the Midlands did not rate sentences significantly differently from each other and the important differences between the groups appear in the differences between northern participants and those from the rest of England. Southern and midland participants were less likely to give as high *feminine* ratings than those from the north. Of the top 10 highest-rated *feminine* sentences, two

164

sentences for the southern participants were given ratings of less than 5, which does not happen for either the midlands or the northern group. Conversely, while the northern group does not have a mean rating for any sentence of less than 4.5, the southern and midland participants rate femininity as low as 3.67 and 3.44 respectively.

It is important to remember that the stimuli provided for this survey were all from a speaker from the north of England and the participants were specifically prompted to rate a Yorkshire accent. This data suggests that southern and midlands participants perceive femininity in Yorkshire speakers significantly differently from their northern counterparts. Further research on this concept would benefit from including speakers with southern accents. This would allow researchers to investigate if southern and midland speakers have generally more restricted views of femininity, or if they are particularly narrow in their views on northern femininity.

These differences in ratings are not as strong when considering the ratings of *homosexual*. While there are still some slight differences between the groups, these differences are not as large as ratings of femininity proved to be. The largest voting range appeared in the midland participants, with a range of 1.11 from the highest to lowest mean rating, while the southern participants had a range of 0.57 and the northern group had a range of 0.51. In the southern and northern groups, participants are less likely to give "agree" or "strongly agree" ratings for homosexual, but these groups will rate "disagree" or "strongly disagree" more often.

As with femininity, this research would benefit from further studies with a southern speaker to test the possibility that, at least for northern listeners, they were less likely to rate a sentence as *homosexual* because it was more similar to how people around them sound, or if they are unlikely to rate any speaker as sounding *homosexual*.

The differences between participants based on sexual orientation, gender, and English regions show that there are nuanced differences in perceptions of traits like femininity and homosexuality. While it may be easy to consider the data in a holistic way, in the way the data is initially presented in the second part of the

results (Section 4.3.2), these closer investigations into groups of participants highlight just how important the listener's identity is, as well as the speaker's, for successful communication.

4.4.4 Quality correlations

The final section of the discussion will consider the qualities that correlated with one another in this study. These correlations indicate that certain qualities are directly related to others, and these may have larger implications for societal views of gender.

4.4.4.1 Perceived femininity

As presented in the results section, there was a significant correlation between F0 alteration and perceived femininity. This positive correlation may not be particularly surprising given the previous research on typical feminine F0 means, as evident in Fraccaro *et al.* (2011) and Munson (2007a). This correlation leads to larger questions when considering the other qualities that increased F0 also correlates with.

Femininity was found to be positively correlated with friendliness, intelligence, and trustworthiness. These results are surprising given that intelligence had a significant negative correlation with F0 alteration, but femininity had a significant positive correlation with F0 alteration. Though these results may appear conflicting, the data suggests that qualities such as friendliness, intelligence, and trustworthiness are judged in connection to femininity and would thus be inherently gendered.

4.4.4.2 Homosexual correlations

The particular interest of the present research is to understand what a gay voice for women might be perceived as, and as part of that, are there certain qualities that correlate with a perceived gay voice?

The results indicate that homosexuality is positively correlated with perceptions of low pitch and speaker age. However, homosexuality is negatively correlated with

femininity. The participants in the present study do not perceive a *homosexual* voice as also being a *feminine* voice, which was also found in Munson (2007a). This may also connect to the negative correlations with *friendly* and *trustworthy*, which positively correlated with *feminine* but negatively correlated with *homosexual*.

4.5 Conclusion

This survey has provided considerable insights into how participants perceive a gay voice for women. While there may not be many stereotypes for a gay voice for women, there are some distinct patterns that emerge from this data that suggest there might be underlying assumptions about a female gay voice. Participants typically rated sentences with lower F0 higher for the quality of *homosexual* and sentences with increased F0 lower for the quality of *homosexual*.

The results are complicated when considering the different groups that took part in the study and when the results are viewed in a more nuanced way. This study has demonstrated the validity of considering the identity of the participants when conducting perceptual research on a gay voice.

However, even when considering the group differences, a pattern still emerges that sentences with lower F0 are typically rated as sounding more *homosexual* than sentences with higher F0.

As with much of the previous perception work, these perceptions were all based on speech samples produced by a cis-gender, white speaker. Further research that includes non-cis speakers or speakers of different ethnicities might indicate that these perceptions are based on certain identity factors that are not applied to others.

Chapter 5: General Discussion and Reflection

5.1 Introduction

The goal of this chapter is to place the two major studies of this thesis into a wider context in two separate ways. The first section will consider both the production and perception studies together in an effort to understand how these interact and what these studies reveal about a gay voice for women. There will also be an exploration of how people identify themselves and the terminology that they use, as this was important in both studies.

The second part of this chapter aims to reflect on how this entire project fits outside of just language and sexuality research. This will constitute a focus on methodology and how to approach linguistic research in the future, and will present arguments for future work.

5.2 General Discussion

While the two studies presented in this thesis were completed separately, using different methodologies and participants, they can be viewed together to further the understanding of how speakers may use voice to index their sexual orientation, and of how listeners might perceive certain phonetic qualities as indicative of sexual orientation.

5.2.1 Extent of a gay voice

The participants in both studies were hesitant about acknowledging the existence of a gay voice for women. First, in the production study, participants were asked if they thought a gay voice for women existed and if it did, what it would sound like. While a few participants suggested that there was a connection to a "laddish" voice or other connections to masculinity, they did not say that they felt a strong sense of there being a gay voice from the participants as a whole. Some said that it was not something they had ever considered before and suggested they would have to rely on stereotypes in order to answer the question. Participants typically said that if

there was a way to identify a woman's sexual orientation without being explicitly told, as in "gaydar", it was more connected to physical appearance (e.g. dress, hairstyle).

In the perception study, a large proportion of the participants answered "neither agree nor disagree" to every sentence stimulus when rating the quality of *homosexuality*. All the other qualities that were included in the survey were rated with some level of agreement (e.g. "strongly disagree" or "agree"), but *homosexual* was not. This suggests either unfamiliarity with a gay voice for women or discomfort in acknowledging what that voice might sound like. I argued in section 4.4.1 that both the unfamiliarity with a stereotyped gay voice for women and the discomfort of acknowledging such a voice account for the non-committal nature of the responses on the *homosexual* rating scale in the survey.

Considering these two studies together, it appears that British people generally do not have strong views about a gay voice for women. Even in considering the discomfort they may have felt acknowledging stereotypical sexual orientation categories, I argued in section 4.4.1 that this discomfort may have been overcome if there were stronger stereotypes about this specific voice. While participants may have felt uncomfortable relying on stereotypes if such strong stereotypes existed, they may have been more inclined to acknowledge them.

This hesitancy in making strong claims about a gay voice for women matches much of the previous research. Cameron and Kulick (2003, p. 86) argue that there is not a coherent lesbian identity, which has meant there is not a clear "lesbian language", in contrast to the more deep-seated stereotypes of a male gay voice. The consensus among much of the previous research (Waksler, 2001; Munson, 2007a; Rendall, Vasey and McKenzie, 2008) is that if there are any significant differences between gay and straight female speakers, they are not consistent. This may present a chicken and egg situation, in which it is difficult to know if there are fewer stereotypes about a gay voice for women because voice is not used to index sexual orientation, or the features are not commonly used because voice is not a commonly stereotyped way of indicating a woman's sexual orientation.

It is also important to acknowledge the reliance on appearance from the production study participants. This focus on female appearance is not a new phenomenon, as Fredrickson and Roberts (1997, p. 174) wrote on female objectification in the 1990s. They write, "[the] common thread running through all forms of sexual objectification is the experience of being treated *as a body* (or collection of body parts) valued predominantly for its use to (or consumption by) others" (emphasis in original work). The authors also acknowledge that this was a commonly accepted view by feminists before their work, and their research was conducted to look at how this objectification happens and its effects on women's mental health.

More current research also acknowledges the frequency of appearance judgements for women, often having greater ramifications than male appearance judgements. In their study of how gender facial cues affect political success, Hehman *et al.* (2014, p. 821) found that "gendered cues uniquely predict female politicians' electoral success beyond these factors, suggesting a discrepancy between traits used to evaluate male and female politicians." Similarly, Banchefsky *et al.*'s (2016) study of the evaluation of feminine appearance and perceptions of engineering qualifications found that appearance was significant in determining if a woman was perceived to be a scientist. They write that their study "contributes to research suggesting that appearance is more valued, scrutinized, and consequential for women than men" (p.107).

While this present study did not focus on dress or physical appearance as cues to female sexual orientation, based on the responses of the participants and previous research it is important to acknowledge that appearance may be significant in how people perceive sexual orientation in women. Voice may not be the principal factor people use when making assumptions about female sexual orientation, but this does not mean that it is not important.

Despite the lack of stereotypes around a gay voice for women, or a potential discomfort in acknowledging it on the part of participants, this study has shown that there are significant features of speech that are both associated with homosexuality and used by gay women in this particular community. While this female gay voice may be less stereotyped than a male gay voice, the present study indicates that there

may be some qualities certain speakers use to index at least a certain type of gay identity.

5.2.2 Self-identification labels

Though the focus of this research was a sociophonetic study of the potential phonetic correlates of a gay voice for women, other topics proved to be of note and worthy of further consideration. One such topic is terminology and how women, particularly young, gay ones, may identify themselves.

As discussed in the production study (section 3.3.1.1), there was an interesting conversation with participants around how they refer to their own sexual orientation. While many participants said that they did not care what they were called, many also stated that they prefer the term "gay" to "lesbian". Some participants used particularly strong language when discussing this terminology and even indicated that they "hated" the word "lesbian". Others said they did not like being boxed in a narrow category and preferred "gay" because it is more generic. None of the production participants used the term "homosexual" to describe themselves, and no one said they preferred it when asked.

There was also variation in the perception study in relation to how women labelled their own sexual orientation when they identified as gay. There were three female participants that identified as gay, but these three all used different terms. One person identified as "lesbian", a second as "homosexual", and finally, a third as "gay". There were also two women who identified themselves as "queer", which could fit within the gay identity but could also be completely separate.

This usage matched the production study, in which gay participants used the word "lesbian" on their questionnaire forms, even when later in the interview they acknowledged that not only did they prefer the term gay, they particularly did not like the word "lesbian". It was argued in Section 3.4.1 that when completing forms or using a more formal register, women may be more inclined to use the word "lesbian" to describe their sexual orientation to avoid confusion. It is possible that

this perception survey is part of the same category as the formal register and participants changed the word they use when identifying their sexual orientation.

Apart from register, one of the potential reasons for this difference in terminology may be age. The participant that identified as "homosexual" in the perception study was in the 35-44 age bracket. The other participants that identified as "lesbian", "gay", or "queer" were all in the 25-34 age bracket. While that does not account for the use of "lesbian" by one of the participants, it may be a sign that "gay" and "queer" are terms used by younger people.

Writing in the early 2000's, Cameron and Kulick (2003, p. 27) assert that "many lesbians prefer the gender-specific term 'lesbian' to 'gay', which, they argue, obscures the presence of women by subsuming them under a label whose primary reference is to men." Though there is not a specific group of speakers this view is attributed to, it is likely that this group of women who prefer the term "lesbian" are older than many of the participants in the present study give that the book was published in 2003. This change of preference from Cameron and Kulick's writing and the present study may support the age-dependent hypothesis about terminology preferences.

In fact, in the more recent study by Sauntson and Morrish (2012), the researchers also found that the participants preferred the term "gay" to "lesbian" and take the stance of using "the participants' preferred term of 'gay' rather than 'lesbian' when discussing the data" (p. 153). This 2012 study also focused on a women's football team and the participants were between 19 and 21 years old. While the participants in the present study are generally older than those in Sauntson and Morrish's research, they appear to have similar stances regarding how to identify their own sexual orientation. Though this is not an exhaustive amount of data, there do appear to be indications that there is a shift in terminology preferences with the younger generation of gay women.

Despite this increase in women's use of the term "gay", it is still predominantly used by men, and they may still be the primary reference. As Blank (2011, p. 134) writes in her etymology of the word "lesbian", "[t]hough we may alternatively call

ourselves 'gay' or 'homosexual,' such terms are, for some, invariably and problematically gendered male." This view of "gay" as male is supported by the perception study responses.

In comparison to the inconsistent terminology used by gay women in the perception study, all of the male participants that identified as gay use the word "gay", with the exception of one participant who wrote "gay, queer". This usage spanned the age brackets, with participants in the 18-24 and 45-54 groups using the same term. This implies that the use of "gay" may still be predominantly associated with men and the preferred way to self-identify for male speakers.

However, the participant who chose the word "queer" to describe himself was the only one of this group to be in the 18-24 age bracket. As noted above, two female participants also identified as queer, and a third non-binary person. All of these participants who identified as queer were under 35 years old. This further confirms that younger people may more frequently use "queer".

While these are only preliminary findings, and obtaining them was not the original goal of the current project, this research has highlighted a potential shift in terminology used to describe sexual orientation, particularly among gay women. Further research into these trends could illuminate why people may prefer certain terms over others, and the context in which they prefer these terms. It would also be telling to see if this shift is a growing trend among young people and "queer" could be the new standard when describing one's sexual orientation.

5.2.3 Gay voice, pitch, and /s/

Pitch was analysed in both studies in order to investigate whether there is any connection between what people do and what others perceive. It was found that mean pitch is significant in both production and perception in these two different studies and for this reason it may be a key factor in a gay voice for women based on this group of speakers. As acknowledged before, this idea of a gay voice is a set of qualities that may index a gay identity, but not necessarily every speaker who is gay would use this particular voice and not everyone that uses this particular voice is

gay. This distinction is important when considering that Scarlett, the voice of the perception study stimuli, does not identify as gay.

While there were significant findings associated with pitch and sexual orientation, Chapter 3 argued that pitch was indicative of a specific gay identity employed by certain participants in the study. Participants that were on the Yorkshire Town Ladies football team had significantly lower mean pitches than participants did that were not on the team. As the team participants were predominantly gay, it is possible that the use of a lower pitch may be connected not just with being gay, but also to being sporty as argued in section 3.5. This sporty identity could account for the significant differences in pitch range and /s/ production according to team association.

In that section, I also argued against identifying this particular sporty group of speakers as "butch", because none of the participants identified herself in this way. However, what was significant with this group was that they were also not adhering to prescribed gender norms and did not consistently present stereotypical feminine traits. For this reason, there may be some loose connections between the sporty identity of the YTL members and butchness, as both challenge hegemonic gender norms for women.

In the perception survey, there was a significant negative correlation between pitch and homosexuality, as well as a significant positive correlation between pitch and femininity. From the data it is possible to see that participants identified lower pitch as being less feminine, as well as being more homosexual. This correlation between pitch and hegemonic gender norms is key when evaluating the significant findings in the present study.

As has been pointed out before by Munson (2007a), the connection between pitch and identity may hinge more on the idea of gender identity than sexual orientation. How participants may have rated homosexuality could in fact have been judgments on voices they felt did not meet hegemonic norms for a female speaker (e.g. female speakers should have higher voices). In this sense, if a voice was not meeting gender norms, then the speaker must be homosexual. For this reason, speakers who

174

are sporty may be more identifiable as gay because they do not adhere to gender norms.

This view of a gay voice working in tandem with gay speakers actively distancing themselves from gendered norms connects with previous research. As Jones (2018, p. 1) writes in her chapter summarising previous research on lesbian identity, "most existing studies of interaction between women identifying as lesbian have shown them to draw on symbols and practices that allow the positioning of themselves as ultimately different from, or even in opposition to, the heteronormative mainstream image of femaleness." The use of a lower mean F0, as well as a smaller F0 range, may be part of this opposition to femaleness, often associated with masculinity.

As Jones (2018, p. 1) asserts, it may be that some gay women are "doing butch identity", which is still a female identity, but one that does not adhere to feminine gender norms. It is possible that the members of the YTL are more frequently doing this sporty identity than the gay participants who are not members of the team. However, this does not mean that every woman that plays football is inherently "sporty" in the sense of this study, or even that every gay woman that plays football is sporty. Since the core of the football team was created by asking friends to join, it is likely that this group of women happen to have more in common than would individuals selected at random, which created a friendship, which ultimately led to a football team. Their relationship, both on and off the pitch, may be due to their shared sporty identity.

While there are indications of areas that could be researched further in connection to a gay voice for women, there would need to be more consideration of how both the production and perception of a gay voice also interact with the concept of femininity and gender norms, and how this is realised in a local community. It is also important to note that voice may be how this particular group of sporty speakers indexes their sexual orientation, but that does not mean that it would be universal, even for other groups that maybe be doing a similar identity associated with sport.

A second consideration for further studies is the significance in /s/ production of participant perception of sexual orientation. While Scarlett's voice was digitally

altered to change the pitch, matching the pitch register of some of the sporty gay participants, her /s/ was not altered in any way. Based on the production study, it is clear that Scarlett has /s/ productions that are more typical of the straight, non-members of this study. In fact, Scarlett's mean /s/ centre of gravity is higher than that of all the gay members of YTL, apart from Elizabeth. This /s/ production could be significant in how participants perceive sexual orientation and therefore would need to be tested for in future work.

The production study demonstrates that there is significant variation even within this fairly small community of gay speakers within the same county of the UK. This variation demonstrates that a gay voice has to inherently be viewed as nuanced and specific to a local community. This study shows that there is not a set of features that every speaker employs or even a set of features that gay speakers use that is completely distinct from straight speakers to perform sexual orientation. Instead, there is nuance to these groups and they may be performing a complex set of identities that intersect in a specific way that has led to the features considered in this study. These features could also vary depending on the context of the recording, such as interlocuter and the topic being discussed.

Based on this discussion of the complexity of this group of speakers and the general discussion of intersectionality from section 2.2.3, this nuanced view of a gay voice should be considered not just for this specific project or for a female gay voice, but for sets of features associated with a gay voice for any gender. It is important to consider how the myriad of intersections of different identities create a specific, local performance. While this intersectional view is common in language, gender, and sexuality work (Levon, 2015), it is important that all those engaging with this research appreciate the nuance the work inherently has.

With that nuance being established, one consideration that could have larger scale implications is the interaction between gender norms and sexuality performances, as has been discussed above. Based on previous work such as Munson (2007b), Jones (2018), and this current research, it may be that certain gay identities are based on distinguishing one from typical gender norms. However, in following this research it would be significant to understand what local performances of gender may be, as

these may not be universal or even span across the UK, and it would be important to consider if different gay participants are performing an identity that moves away from typical gender norms based on other identity factors.

5.3 Reflection

While the section above placed the present studies in a wider context of research on the speech of gay and straight women, the purpose of this reflection is to consider how this thesis fits with the previous body of linguistic research on language and sexuality based on methodological approaches. As part of this reflection, it is important to first place the researcher within the project and to investigate the potential ramifications of who is doing the work. Secondly, this section will demonstrate how sexual orientation should be considered further within variationist work, and, on a larger scale, how learning from queer linguistic research can benefit variationist research.

5.3.1 Placing the researcher

As is common in sociolinguistic work, the researcher does not always match the demographic of the participants being studied. It has been acknowledged in previous work that an interviewer may have an influence on the interviewee and a level of accommodation may be present in the data. Mendoza-Denton (2004, p. 479) writes "[the] idea that the *researcher's* identity and ideological positioning vis-à-vis the interviewee crucially contributes to the patterning of data deserves more systematic exploration" (emphasis in original work).

While sexual orientation may not be as strong a linguistic factor as others, given the lack of female gay voice stereotypes acknowledged above, it is possible that some level of accommodation took place in the interviews. Most of the participants knew of my sexual orientation before the interview even started, because of my previous contact with the team. However, in most of the interviews, I did also acknowledge my own sexual orientation, by mentioning my male partner or referring to myself as straight, as means of identifying myself to the participants. Due to the inconsistency of previous knowledge, it is not possible to consider in the present data. Although,

it would be of interest in future research to see if there are any subtle differences in a speaker's phonetic behaviour after having "outed" myself to the participants.

However, other than acknowledging the potential for accommodation, this section is also intended to acknowledge the responsibility of a researcher working with a group of participants that are a potentially vulnerable group. While the UK is one of the more progressive countries when it comes to LGBTQ rights, there are still daily injustices faced by this community and there is still significant progress left to make before reaching any sort of equality. At the time of writing this thesis, there are on-going protests against LGBTQ-inclusive teaching in schools in Birmingham (Kotecha, 2019) and a Member of Parliament report encouraging heath care providers to consider patient sexual orientation, as LGBTQ "people are often less healthy than the wider population, but receive lower levels of care" (*BBC News*, 2019).

As a straight, cis-gender person conducting research with people in the LGBTQ community, I was often concerned by the outcome of this study and the potential to reinforce harmful stereotypes about those involved. While this would never be my intention, I was worried that my lack of personal experience of the injustices LGBTQ people face would blind me to potential pitfalls of the research. Due to my sexual orientation and gender identity, I am inherently coming from a place of privilege when considering LGBTQ issues. I did not want this privilege to blind me to harmful decisions and assumptions I might have brought to the research.

Yet I believe it was this awareness of my privilege and the concern I had for its possible consequences that made me a stronger and more ethical researcher. In a personal communication, a colleague acknowledged "as a queer linguist and a linguist who happens to be queer I'm okay with you doing this work" precisely because I was conscious of my place of privilege when undertaking this project.

This is not an argument against people outside a community conducting research on it or an assertion that only someone within a community can conduct valid research on its members. This is instead an argument that one should be conscious of privileges and advantages that may be in place when conducting research. By at

least being aware of the privileges, one can be more sensitive when approaching participants, how researchers interact with them, and the potential ramifications of the research output of the project. Awareness and critically considering one's own impact on the research allow for better research, both empirically and ethically.

This awareness of privilege does not only fall within queer linguistics or work within the LGBTQ community. It is important to be aware of the position a researcher is in when conducting the work and how that may influence their understanding of the data or even how they approach the participants. This does not stop interesting research from taking place or hold people back, but simply argues that researchers should be as critical of their place within the study as they are with the data they analyse.

5.3.2 Connecting variationist and queer linguistics

Along with considering my own place a straight, cis-gender woman conducting research within an LGBTQ community, another element of this research that had to be reflected upon was interaction between queer linguistics and variationist linguistics. These different viewpoints were borne in mind throughout the present research and an effort was continually made to meet the goal of bringing the two together in this project. In working on this balance between variationist linguistics and queer linguistics, I ended up following a more third wave variationist approach, discussed in section 2.2.1.

One common methodological approach to both queer linguistics and third wave variationist linguistics is ethnography. As discussed in the methodology in section 3.2.2, this project was initially set to follow a community of practice (Eckert and McConnell-Ginet, 1992) and investigate how language intersected not only with sexual orientation, but also with team association, gender identity, and other identities that were relevant to the group. At the beginning of the project, the team had gay and straight players, as well as some members who had been on the team since its inception and others who had only been a member for a season. This would have allowed for in-depth analysis of how all of these factors interacted and might have been represented in language use. This original project would have been more

in line with previous language and sexuality research that follows a more ethnographic approach, such as Podesva (2011), Levon (Levon, 2011), Jones (2012), Zimman (2017), and many more.

However, due to the unforeseen circumstances of the team disbanding, this project had to be recast around the recordings that had already been acquired. This recasting of the project could have led to a more first or second wave variationist approach, as dictated by the data that was available. In some ways these variationist approaches are at odds with what Eckert and Podesva (2011, p. 8) assert when they write: "we suggest first that categories be ethnographically significant and second that they be viewed as products of, rather than explanations for, variation patterns." Instead of reverting back to more traditional first or second wave approaches based on the unforeseen setbacks, this study was able to maintain a third wave approach by how the data was analysed.

One of the difficult elements when doing a variationist project with an orientation towards queer linguistics was working with categories, and how to categorise participants without relying on stereotypes. One of the significant missions of queer linguistics is work against essentialism. Indeed, that is why *queer* was initially envisioned to be a "non-signifier signifier" that does not particularly index any one identity (Barrett, 2002). It was thought that by having this openness and shift away from categories, there would not be a risk of stereotyping and overgeneralising.

However, in variationist work, there is inherently a need to create categories that speakers fall within in order to compare different groups. These categories can be difficult and uncomfortable. The solution that the present research took based on a third wave approach was to allow participants to categorise themselves as much as possible. Questionnaires and surveys were intentionally designed to have open spaces that participants could use to identify themselves with whatever terminology they saw fit, in the amount of detail they wanted to include. By having these open-ended questions, categories could be made around the participants, instead of the other way around.

It is important to acknowledge that in some instances, particularly when creating a secondary group of speakers for the production study, there were certain demographic traits that were being sought. However, these traits were acknowledged by either the participants themselves or the acquaintance that acted as an intermediary between the researcher and the participant. There was nevertheless still a place for participants to identify themselves and the research would have accommodated to the different identity categories that were acknowledged.

A second factor to consider in relation to the identity categories of the different participants was to avoid combining too many groups together in a way that would obscure how an individual identified. While at times this was necessary, as highlighted in the perception study, allowing for multiple identities in the analysis would permit richer and deeper understanding of the data. For example, in the perception study, participants were not all considered as a homogenous group labelled "English" and left at that, but instead there was an attempt to go even narrower and to consider the region they were from.

This openness to different identity categories allows third wave variationist work to fit under the umbrella of queer linguistics. While other queer linguistic research may not identify the group as clearly as variationist work typically has to, there can still be an interaction between the two fields of study.

Not only does this study show how variationist research can fit within queer linguistics, it also demonstrates the necessity of considering speakers' and listeners' sexual orientation when conducting variationist work. Outside of studies that are particularly interested in sexual orientation, the sexual orientation of the participants is rarely considered. However, as this study shows there are some subtle differences between the gay and straight participants. This has also been seen in previous research that has demonstrated differences in both production and perception in gay and straight voices. Just as participant gender and ethnicity are often cited in variationist research, it seems clear that sexual orientation should be a factor that is also considered in future work.

This third wave variationist approach has also shown the significance locally significant communities has on data and why local communities should be considered more frequently in more sociolinguistic data. If there was a reliance on macro-social categories such as gay and straight, it would not have been possible to see the subtle variation between the group of gay speakers. This subtle variation does not only exist within this specific gay community, but exists within all macro-social categories. If sociolinguists truly want to understand the interaction between identity and linguistic variation, then it is key that there is nuance in methodology and analysis that understands how different people of similar macro-social categories may perform specific identities based on their unique intersections with other identities.

This reflection has sought to discuss how sociolinguistic work can continue to expand and develop further. While this thesis may be comprised of only two studies, it has presented an opportunity to critically consider methodology, and has allowed for a continuous discussion on how best to collect linguistic data.

5.4 Conclusion

This chapter considered the broader implications of this research outside one particular study. First, there was a consideration of how the two studies interacted with each other and what the similarities and differences might signify about a gay voice for women. Second, this research as a whole was placed in a larger linguistic context, first via reflecting upon how the researcher influences the work, and then how the present work can contribute to future research in this area. While there are many directions future work can take based on this initial study, there are important lessons that came from this thesis.

6. Conclusion

This thesis has demonstrated that there is scope for further research when considering female sexual orientation and phonetic features. Mean F0 was significantly different between gay and straight participants in the production study, and was also significant in how listeners perceived a speaker in the perception study. When considered alone, realisations of /s/ were also significantly different for gay and straight speakers.

However, as the thesis has demonstrated, these differences were more nuanced than simply a gay and straight division. It became apparent that there was a particular "sporty" identity that gay members of the Yorkshire Town Ladies football team were presenting that was different from the identity claimed by the other gay speakers or the other straight team member. This thesis has argued that the participants were presenting a more specific type of identity that intersected with sexual orientation, and that this identity predisposes speakers to behave in particular ways with respect to patterns of phonetic variation in their speech.

The identity of the gay YTL members intersected with gender identity, and this thesis argued that by being both openly gay and actively part of a football team, they moved further way from a normative gender identity. This non-normative gender identity may also connect with what listeners perceived as sounding more *homosexual*. As was seen in previous work by Munson (2007a), listeners appear to be sensitive to gender norms, and their ratings for sexual orientation are influenced by associations with gender normativity.

Along with the phonetic features that were analysed, interesting qualitative data emerged from the interview with the participants and in particular the gay participants in the production study presented in Chapter 3. Through the interviews there was a clear acknowledgment that many did not believe a gay voice for women existed at all, even if they believed such a thing existed for men. There were also signs of shifts in terminology preferences by the gay players, who predominantlypreferred the term *gay* over *lesbian*. Both the quantitative and qualitative findings allow many avenues for future work.

6.1 Future research

In Chapter 3 there was a discussion of the changing dynamic of the football team and its ultimate demise due to lack of participation. Further research would benefit from following a more stable community in order to go into more depth with their views on their own gender identity and sexual orientation, and to consider further how phonetic features may be influenced by different gender identities. Other sports teams may provide fascinating data based on the age of the participants and the type of sport involved; would gay gymnasts (a sport more gender-normative than football) have different phonetic patterns from those of the YTL members? Similarly, could people who consider themselves LGBTQ activists have more non-normative gender features. There is scope for considering how different communities interact with gender and sexual orientation.

Similarly, there is a huge penitential in language and sexuality research for including participants that are not white or middle-class. While this study did consider social class and there were different self-identified social classes present, it would be beneficial to forefront this identity further and to expand the background questions in order to place speakers more decisively in social class categories.

The perception study would benefit from including stimuli produced by a speaker from the South of England in order to investigate whether regional accents are influential in how listeners perceive qualities like femininity and homosexuality. Based on the different ratings given by participants in different regions, it appears this may be the case and is worth exploring further.

This thesis has asked the question "can a woman sound gay?" and the answer, as expected from the outset, is "it is complicated". While there were some significant differences between the speech of gay and straight participants, and listeners significantly rated stimuli with a lower F0 as sounding more *homosexual*, these qualities also intersect with gender identity and geographic region. As sociolinguistic researchers know, as well as all those who work in the social

sciences, people are complicated, with an infinite number of layers to make up how they identify and what is significant in any given context. This thesis has taken an in-depth view of a small number of these layers in an attempt to understand how speakers may present themselves to the world.

References

Ainsworth, C. (2015) 'Sex redefined', *Nature*, 518(7539), pp. 288–291. doi: 10.1038/518288a.

AudacityTeam (2019) 'Audacity'.

Banchefsky, S. *et al.* (2016) 'But You Don't Look Like A Scientist!: Women Scientists with Feminine Appearance are Deemed Less Likely to be Scientists', *Sex Roles*. Sex Roles, 75(3–4), pp. 95–109. doi: 10.1007/s11199-016-0586-1.

Barrett, R. (1997) 'The "Homo-genius" Speech Community', in Livia, A. and Hall, K. (eds) *Queerly Phrased*. Oxford: Oxford University Press, pp. 181–201.

Barrett, R. (2002) 'Is Queer Theory Important for Sociolinguistic Theory?', in Campbell-Kibler, K. et al. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford: CSLI Publications, pp. 25–44.

Barrett, R. (2014) 'The Emergence of the Unmarked: Queer theory, ideology, and formal linguistics', in Zimman, L., Davis, J. L., and Raclaw, J. (eds) *Queer Excursions: Retheorizing binaries in language, gender, and sexuality*. Oxford: Oxford University Press, pp. 195–223.

BBC News (2019) '"Ask about sexual orientation to improve LGBT inequalities"', 22 October.

Beal, J. C. (2010) *An Introduction to Regional Englishes*. Edinburgh. Edinburgh.

Berlant, L. and Warner, M. (1995) 'What Does Queer Theory Teach Us about Intersectionality?', *PMLA*, 110(3), pp. 343–349.

Berryman-Fink, C. L. and Wilcox, J. R. (1983) 'A multivariate investigation of perceptual attributions concerning gender appropriateness in language', *Sex Roles*, 9(6), pp. 663–681. doi: 10.1007/BF00289796.

Bland, L. (1996) 'The humorous side of butch and femme', in Warner, N. et al. (eds) *Gender and belief systems: Proceedings of the fourth Berkeley Women and Language Conference*. Berkeley, CA, pp. 73–84.

Blank, P. (2011) 'The proverbial "Lesbian": Queering etymology in contemporary critical practice', *Modern Philology*, 109(1), pp. 108–134. doi: 10.1086/661977.

Boersma, P. and Weenink, D. (2019) 'Praat: doing phonetics by computer'.

Van Borsel, J., Vandaele, J. and Corthals, P. (2013) 'Pitch and pitch variation in lesbian women', *Journal of Voice*. Elsevier Ltd, 27(5), pp. 656.e13-656.e16. doi: 10.1016/j.jvoice.2013.04.008.

Brontsema, R. (2004) 'A Queer Revolution: Reconceptualizing the Debate Over Linguistic Reclamation', *Colorado Research in Linguistics*, 17(1), p. 9.

Brown, M. *et al.* (2000) 'Pitch change in male-to-female transsexuals: has phonosurgery a role to play?', *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, 35(1), pp. 129–136. doi: 10.1080/136828200247296.

Brown, P. and Levinson, S. (1987) *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Butler, J. (1990) *Gender Trouble Feminism and the Subversion of Identity*. New York: Routledge.

Cameron, D. and Kulick, D. (2003) *Language and Sexuality*. Cambridge: Cambridge University Press.

Campbell-Kibler, K. *et al.* (eds) (2002) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford.

Campbell-Kibler, K. (2007) 'Accent, (ING), and the social logic of listener perceptions', *American Speech*, 82(1), pp. 32–64. doi: 10.1215/00031283-2007-002.

Caudwell, J. (2007) 'Queering the Field? The complexities of sexuality within a lesbian-identified football team in England', *Gender, Place & Culture*, 14(2), pp. 183–196. doi: 10.1080/09663690701213750.

Cave, A. (2001) *Language variety and communicative style as local and subcultural identity in a South Yorkshire coalmining community*. University of Sheffield.

Chen, M. Y.-C. (1998) '"I am an Animal!": Lexical Reappropriation, Performativity, and Queer', in *Engendering Communication: Proceedings from the Fifth Berkeley Women and Language Conference*, pp. 128–140.

Chesebro, J. W. (ed.) (1981) *Gayspeak: Gay Male & Lesbian Communication*. New York: The Pilgrim Press.

Chun, C. W. (2019) 'Language, discourse, and class: What's next for sociolinguistics?', *Journal of Sociolinguistics*, 4(23), pp. 332–345. doi: 10.1111/josl.12359.

Cooper, P. (2015) 'Enregisterment in historical contexts: nineteenth century Yorkshire dialect', *Dialectologia*, 14, pp. 1–16.

Cooper, P. (2019) 'The enregisterment of "Barnsley" dialect: Vowel fronting and being "broad" in Yorkshire dialects', *Language and Communication*. Elsevier Ltd, 64, pp. 68–80. doi: 10.1016/j.langcom.2018.10.001.

Crenshaw, K. (1989) 'Demarginalizing the intersection of race and sex: a black feminist

critique of antidiscrimination doctrine, feminist theory and antiracist politics',
University of Chicago Leago Forum, pp. 139–68.

Crist, S. (1997) 'Duration of onset consonants in gay male stereotyped speech', *University of Pennsylvania working papers in linguistics*, 4(3), pp. 53–70.

Cruttenden, A. (1986) *Intonation*. Cambridge University Press.

Eckert, P. (2000) *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.

Eckert, P. (2012) 'Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation', *Annual Review of Anthropology*, 41(1), pp. 87–100. doi: 10.1146/annurev-anthro-092611-145828.

Eckert, P. and McConnell-Ginet, S. (1992) 'Think Practically and Look Locally: Language and Gender as Community-Based Practice', *Annual Review of Anthropology*, 21(1), pp. 461–490. doi: 10.1146/annurev.anthro.21.1.461.

Esling, J. H. and Edmondson, J. A. (2010) 'Acoustical analysis of voice quality for sociophonetic purposes', in Di Paolo, M. and Yaeger-Dror, M. (eds) *Sociophonetics: A student's guide*. London: Routledge, pp. 131–148.

Fecher, N. (2011) 'MeasureSpectralMomentsCreateTableExcel_Nat2'. Department of Language and Lingustics, University of York.

Fecher, N. (2015) 'Praat pitch alteration script'. Department of Language and Lingustics, University of York.

Feinberg, D. R. *et al.* (2008) 'The role of femininity and averageness of voice pitch in aesthetic judgments of women ' s voices', *Perception*, 37, pp. 615–624. doi: 10.1068/p5514.

Finnegan, K. (2015) 'Sheffield', in Hickey, R. (ed.) *Researching Northern English*. John Benjamins, pp. 227–250.

Flanders, C. E. *et al.* (2017) 'Defining Bisexuality: Young Bisexual and Pansexual People's Voices', *Journal of Bisexuality*. Taylor & Francis, 17(1), pp. 39–57. doi: 10.1080/15299716.2016.1227016.

Fletcher, T. (2012) '"All Yorkshiremen are from Yorkshire, but some are more 'Yorkshire' than others": British Asians and the myths of Yorkshire cricket', *Sport in Society*, 15(2), pp. 227–245. doi: 10.1080/17430437.2012.637735.

Foucault, M. (1978) *The History of Sexuality Volume 1: An Introduction*. Edited by Trans. Robert Hurley. New York: Random House.

Fraccaro, P. J. *et al.* (2011) 'Experimental evidence that women speak in a higher voice

pitch to men they find attractive', *Journal of Evolutionary Psychology*, 9(1), pp. 57–67. doi: 10.1556/JEP.9.2011.33.1.

Fredrickson, B. L. and Roberts, T. A. (1997) 'Objectification Theory: Toward understanding women's lived experiences and mental health risks', *Psychology of Women Quarterly*, 21(2), pp. 173–206. doi: 10.1111/j.1471-6402.1997.tb00108.x.

Gaudio, R. P. (1994) 'Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men', *American Speech*, 69(1), pp. 30–57. doi: 10.2307/455948.

Gelfer, M. P. and Schofield, J. (2000) 'Comparison of Acoustic and Perceptual Measures of Voice in Male-to-Female Transsexuals Perceived as Female Versus Those Perceived as Male', *Journal of Voice*, 14(1), pp. 22–33.

Giles, H. (1970) 'Evaluative reactions to accents', *Educational Review*, 22(3), pp. 211–227. doi: 10.1080/0013191700220301.

Grabe, E., Post, B. and Nolan, F. (2001) 'The IViE Corpus'. Department of Linguistics, University of Oxford.

Haddican, B. *et al.* (2013) 'Interaction of social and linguistic constraints on two vowel changes in northern England', *Language Variation and Change*, 25(3), pp. 371–403. doi: 10.1017/S0954394513000197.

Hall, K. (2013) 'Commentary I: 'It's a hijra'! Queer linguistics revisited', *Discourse and Society*, 24(5), pp. 634–642. doi: 10.1177/0957926513490321.

Hall, K., Levon, E. and Milani, T. M. (2019) 'Navigating normativities: Gender and sexuality in text and talk', *Language in Society*, 48(4), pp. 481–489. doi: 10.1017/S0047404519000447.

Hazenberg, E. (2012) *Langauge and identity practice: A sociolinguistic study of gender in Ottawa, Ontario*. St. Johns, NL: Memorial University of Newfoundland.

Hehman, E. *et al.* (2014) 'Early Processing of Gendered Facial Cues Predicts the Electoral Success of Female Politicians', *Social Psychological and Personality Science*, 5(7), pp. 815–824. doi: 10.1177/1948550614534701.

Hellmuth, S. and Almbark, R. (2017) 'Intonational variation in Arabic Corpus 2011-2017'. Colchester, Essex: UK Data Archive.

Hiraga, Y. (2005) 'British attitudes towards six varieties of English in the USA and Britain', *World Englishes*, 24(3), pp. 289–308. doi: 10.1111/j.0883-2919.2005.00411.x.

Holmes-Elliott, S. and Levon, E. (2017) 'The substance of style : Gender , social class and interactional stance in / s / -fronting in southeast England', *Linguistics*, 55(5), pp. 1045–1072. doi: 10.1515/ling-2017-0020.

Honorof, D. N. and Whalen, D. H. (2005) 'Perception of pitch location within a speaker's F0 range', *The Journal of the Acoustical Society of America*, 117(4), pp. 2193–2200. doi: 10.1121/1.1841751.

Johnstone, B. (2016) 'Enregisterment: How linguistic items become linked with ways of speaking', *Language and Linguistics Compass*, 10(11), pp. 632–643. doi: 10.1111/lnc3.12210.

Jones, L. (2011) '"The only dykey one": constructions of (in)authenticity in a lesbian community of practice.', *Journal of homosexuality*, 58(6–7), pp. 719–41. doi: 10.1080/00918369.2011.581917.

Jones, L. (2012) *Dyke/Girl : Language and Identities in a Lesbian Group*. Palgrave Macmillan.

Jones, L. (2018) 'Lesbian Identity Construction', in Hall, K. and Barrett, R. (eds) *The Oxford Handbook of Language and Sexuality*. Oxford: Oxford University Press, pp. 1–23. doi: 10.1093/oxfordhb/9780190212926.013.28.

Jones, M. J. (2002) 'The origin of definite article reduction in northern English dialects: Evidence from dialect allomorphy', *English Language and Linguistics*. University of York, 6(2), pp. 325–345. doi: 10.1017/S1360674302000266.

Kiesling, S. F. (2004) 'Dude DUDE', *American Speech*, 79(3), pp. 281–305.

Kiesling, S. F. (2009) 'Style as Stance', *Stance: Sociolinguistic Perspectives*, (1966), pp. 171–194. doi: 10.1093/acprof:oso/9780195331646.003.0008.

Kirsch, M. H. (2000) *Queer Theory and Social change*. London: Routledge.

Klofstad, C. A., Anderson, R. C. and Peters, S. (2012) 'Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women', *Proceedings of the Royal Society B: Biological Sciences*, 279(1738), pp. 2698–2704. doi: 10.1098/rspb.2012.0311.

Kotecha, S. (2019) 'LGBT teaching row: Government issues advice on handling school protests', *BBC News*, 9 October.

Kulick, D. (2000) 'Gay and lesbian language', *Annual review of anthropology*, 29(2000), pp. 243–285. doi: 10.1146/annurev.anthro.29.1.243.

Kulick, D. (2002) 'Queer Linguistics?', in Campbell-Kibler, K. et al. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford: CSLI Publications, pp. 65–68.

Labov, W. (1966) *The social stratification of English in New York City*. Washington D.C.Labov, W. (1966) The social stratification of English in New York City.

Washington D.C.: Center for Applied Lingustics.: Center for Applied Lingustics.

Ladefoged, P. (2003) *Phonetic Data Analysis: An Introduction to Fieldword and Instrumental Techniques*. Oxford: Blackwell Publishing.

Ladefoged, P. and Johnson, K. (2011) *A Course in Phonetics*. 6th edn. Wadsworth Cengage Learning.

Lakoff, R. (1973) 'Language and women's place', *Language in Society*, 2, pp. 45–79.

Leap, W. L. (ed.) (1995) *Beyond the Lavender Lexicon: Authenticity, Imagination, and Appropriation in Lesbian and Gay Languages*. Routledge.

Leap, W. L. (2002) 'Not Entirely in Support of Queer Linguistics', in Campbell-Kibler, K. et al. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford: CSLI Publications, pp. 45–64.

Lennes, M. *et al.* (2016) 'Comparing pitch distributions using praat and r', *Phonetician*, 111, pp. 35–53.

Lennes, M. (2016) 'pitch-distributions'. Zenodo. Available at: http://doi.org/10.5281/zenodo.45868.

Levitt, H. M. and Horne, S. G. (2002) 'Explorations of Lesbian-Queer Genders', *Journal of lesbian studies*, 6(2), pp. 25–39. doi: 10.1300/J155v06n02.

Levon, E. (2006) 'Hearing "gay": Prosody, interpretation, and the affective judgments of men's speech', *American Speech*, 81(1), pp. 56–78.

Levon, E. (2011) 'Teasing apart to bring together: Gender and sexuality in variationist research', *American Speech*, 86(1), pp. 69–84. doi: 10.1215/00031283-1277519.

Levon, E. (2015) 'Integrating Intersectionality in Language, Gender, and Sexuality Research', *Language and Linguistics Compass*, 9(7), pp. 295–308. doi: 10.1111/lnc3.12147.

Levon, E. (2016) 'Conflicted Selves: Language, Religion, and Same-Sex Desire in Israel', in Levon, E. and Mendes, R. B. (eds) *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*. Oxford: Oxford University Press, pp. 215–240.

Lick, D. J. and Johnson, K. L. (2016) 'Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments.', *Journal of Personality and Social Psychology*, 110(6), pp. 801–817. doi: 10.1037/pspa0000052.

Lips, H. M. (2005) *Sex & Gender: An Introduction*. 5th edn. London: McGraw Hill.

Livia, A. (2002) 'The Future of Queer Linguistics', in Campbell-Kibler, K. et al. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford: CSLI Publications, pp. 87–98.

Livia, A. and Hall, K. (eds) (1997) *Queerly Phrased: Language, Gender, and Sexuality*. Oxford: Oxford University Press.

Llamas, C. (2015) 'Middlesbrough', in Hickey, R. (ed.) *Researching Northern English*. John Benjamins, pp. 251–270.

Mack, S. and Munson, B. (2012) 'The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the "gay lisp" stereotype', *Journal of Phonetics*. Elsevier, 40(1), pp. 198–212. doi: 10.1016/j.wocn.2011.10.002.

Maegaard, M. and Pharao, N. (2016) '/s/ Variation and Perceptions of Male Sexuality in Denmark', in Levon, E. and Mendes, R. B. (eds) *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*. Oxford: Oxford University Press, pp. 88–104.

McAleer, P., Todorov, A. and Belin, P. (2014) 'How do you say "hello"? Personality impressions from brief novel voices', *PLoS ONE*, 9(3), pp. 1–9. doi: 10.1371/journal.pone.0090779.

Mendoza, N. (2004) 'Language and Identity', in Chambers, J. ., Trudgill, P., and Schilling-Estes, N. (eds) *The Handbook of Language Variation and Change*. Oxford: Blackwell Publishing, pp. 475–499.

Milani, T. M. (2013) 'Are "queers" really "queer"? Language, identity and same-sex desire in a South African online community', *Discourse and Society*, 24(5), pp. 615–633. doi: 10.1177/0957926513486168.

Milani, T. M. (2015) 'Introduction Language and Masculinities...20 Years Later', in Milani, T. M. (ed.) *Language and Masculinities: Performance, Intersections, Dislocations*. London: Routledge.

Mileva, M. *et al.* (2019) 'The Role of Face and Voice Cues in Predicting the Outcome of Student Representative Elections', *Personality and Social Psychology Bulletin*. doi: 10.1177/0146167219867965.

Milroy, L. (1980) *Language and social networks*. Blackwell Publishing.

Moonwomon-Baird, B. (1997) 'Towards a Study of Lesbian Speech', in Livia, A. and Hall, K. (eds) *Queerly Phrased*. Oxford: Oxford University Press, pp. 202–213.

Moore, E. (2004) 'Sociolinguistic Style: A Multidimensional Resource for Shared Identity Creation', *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 49(3), pp. 375–396. doi: 10.1353/cjl.2006.0006.

Moore, E. *et al.* (2009) 'Style, indexicality, and the social meaning of tag questions', *Language in Society*, 38(4), pp. 447–485. doi: 10.1017/S0047404509990224.

Moore, E. and Podesva, R. J. (2009) 'Style, indexicality, and the social meaning of tag questions', *Language in Society*, 38(4), pp. 447–485. doi: 10.1017/S0047404509990224.

Morris, J. (2017) 'Sociophonetic variation in a long-term language contact situation: /l/-darkening in Welsh-English bilingual speech', *Journal of Sociolinguistics*, 21(2), pp. 183–207. doi: 10.1111/josl.12231.

Motschenbacher, H. (2011) 'Taking Queer Linguistics further: Sociolinguistics and critical heteronormativity research', *International Journal of the Sociology of Language*, 212(212), pp. 149–179. doi: 10.1515/IJSL.2011.050.

Motschenbacher, H. and Stegu, M. (2013) 'Queer Linguistic approaches to discourse', *Discourse and Society*, 24(5), pp. 519–535. doi: 10.1177/0957926513486069.

Munson, B. (2007a) 'The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation', *Language and Speech*, 50(1), pp. 125–142. doi: 10.1177/00238309070500010601.

Munson, B. (2007b) 'The acoustic correlates of perceived masculinity, percieved femininity, and percieved sexual orientation', *Language and Speech*, 50(1), pp. 125–142.

Munson, B. and Babel, M. (2007) 'Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech', *Language and Linguistics Compass*, 1(5), pp. 416–449. doi: 10.1111/j.1749-818X.2007.00028.x.

Munson, B., Jefferson, S. V and McDonald, E. C. (2006) 'The influence of perceived sexual orientation on fricative identification.', *The Journal of the Acoustical Society of America*, 119(4), pp. 2427–2437. doi: 10.1121/1.2173521.

Nelson, T. E., Acker, M. and Manis, M. (1996) 'Irrepressible stereotypes', *Journal of Experimental Social Psychology*, 32(1), pp. 13–38. doi: 10.1006/jesp.1996.0002.

O'Connor, J. J. M. and Barclay, P. (2017) 'The influence of voice pitch on perceptions of trustworthiness across social contexts', *Evolution and Human Behavior*. Elsevier Inc., 38(4), pp. 506–512. doi: 10.1016/j.evolhumbehav.2017.03.001.

Ochs, E. (1991) 'Indexing gender', in Duranti, A. and Goodwin, C. (eds) *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press, pp. 335–358.

Partridge, E. (1967) *A Dictionary of Slang and Unconventional English: Colloquialisms and Catch-phrases Solecisms and Catachreses Nicknames Vulgarisms and such Americanisms as have been naturalized*. 6th edn. London: Routledge & Kegan Paul

LTD.

Pierrehumbert, J. B. *et al.* (2004) 'The influence of sexual orientation on vowel production (L)', *The Journal of the Acoustical Society of America*, 116(4), pp. 1905–1908. doi: 10.1121/1.1788729.

Podesva, R. J. (2007) 'Phonation type as a stylistic variable: The use of falsetto in constructing a persona', *Journal of Sociolinguistics*, 11(4), pp. 478–504. doi: 10.1111/j.1467-9841.2007.00334.x.

Podesva, R. J. (2011) 'The California Vowel Shift and Gay Identity', *American Speech*, 86(1), pp. 32–51. doi: 10.1215/00031283-1277501.

Podesva, R. J. and Eckert, P. (2011) 'Sociophonetics and Sexuality: Toward a Symbiosis of Sociolinguistics and Laboratory Phonology', *American Speech*, 86(1), pp. 6–14.

Podesva, R. J. and Van Hofwegen, J. (2014) 'How Conservatism and Normative Gender Constrain Variation in Inland California: The Case of /s/', *University of Pennsylvania Working Papers in Linguistics*, 20(2), pp. 128–137. Available at: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1820&context=pwpl.

Podesva, R. J. and Van Hofwegen, J. (2016) '/s/exuality in Smalltown California: Gender Normativity and the Acoustic Realization of /s/', in Levon, E. and Mendes, R. B. (eds) *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*. Oxford: Oxford University Press, pp. 168–188.

Queen, R. M. (1998) '"Stay queer!" "Never fear!": building queer social networks', *World Englishes*, 17(2), pp. 203–214. doi: 10.1111/1467-971X.00094.

Queen, R. M. (2002) 'A Matter of Interpretation: The "Future" of "Queer Linguistics"', in Campbell-Kibler, K. et al. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. Stanford: CSLI Publications, pp. 69–86.

Queen, R. M. (2006) 'Heterosexism and/in Language', in Brown, E. K. (ed.) *Encyclopedia of Language and Linguistics Volume 5*. 2nd edn. Amsterdam: Elsevier, pp. 289–292.

'queer, adj.3' (2018) *Oxford University Press*. Oxford University Press. Available at: www.oed.com/view/Entry/156236 (Accessed: 16 November 2018).

Rampton, B. (2000) 'Speech community', in Verschueren, J. et al. (eds) *Handbook of Pragmatics*. Amsterdam: John Benjamins, pp. 1–34.

RCoreTeam (2008) 'R: A language and environment for statistical computing'. Vienna, Austira: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.

Rendall, D., Vasey, P. L. and McKenzie, J. (2008) 'The Queen's English: An Alternative,

Biosocial Hypothesis for the Distinctive Features of "Gay Speech"', *Archives of Sexual Behavior*, 37(1), pp. 188–204. doi: 10.1007/s10508-007-9269-x.

Rosenfelder, I. *et al.* (2014) 'FAVE (Forced Alignment and Vowel Extraction) Suite Version 1.1.3'. Available at: https://doi.org/10.5281/zenodo.9846.

Saigusa, J. (2016) 'Jane Lynch and /s/: The Effect of Addressee Sexuality on Fricative Realization', *Lifespans and Styles*, 2(1), p. 10. doi: 10.2218/ls.v2i1.2016.1426.

Sauntson, H. and Morrish, L. (2012) 'How gay is football this year ? Identity and intersubjectivity in a women ' s sports team', *Journal of Language and Sexuality*, 1(2), pp. 151–178. doi: 10.1075/jls.1.2.02sau.

Schwartz, M. F. (1968) 'Identification of Speaker Sex from Isolated, Voiceless Fricatives', *JASA*, 43(5), pp. 1178–1179.

Seals, C. A. (2016) 'The constrastive use of humor by a lesbian comedian for LGBT and general audiences *', *De Genere Journal 2*, 1, pp. 95–107.

Shrikant, N. (2014) '"It's like, 'I've never met a lesbian before!'": Personal narratives and the construction of diverse female identities in a lesbian counterpublic', *Pragmatics*, 24(4), pp. 799–818. doi: 10.1075/prag.24.4.06shr.

Simpson, A. P. (2009) 'Phonetic differences between male and female speech', *Linguistics and Language Compass*, 3(2), pp. 621–640. doi: 10.1111/j.1749-818X.2009.00125.x.

Smyth, R., Jacobs, G. and Rogers, H. (2003) 'Male voices and perceived sexual orientation: An experimental and theoretical approach', *Language in Society*, 32(03), pp. 329–350. doi: 10.1017/S0047404503323024.

Smyth, R. and Rogers, H. (2003) 'Phonetics, Gender, and Sexual Orientation', in *The 2000 Meeting of the Canadian Linguistic Association*. Toronto, pp. 299–311.

Snell, J. (2010) 'From sociolinguistic variation to socially strategic stylisation', *Journal of Sociolinguistics*, 14(5), pp. 630–656. doi: 10.1111/j.1467-9841.2010.00457.x.

Snell, J. (2018) 'Solidarity, stance, and class identities', *Language in Society*, 47(5), pp. 665–691. doi: 10.1017/S0047404518000970.

Strand, E. A. and Johnson, K. (1996) 'Gradient and Visual Speaker Normalization in the Perception of Fricatives', *Natural Langauge Speech Processing and Speech Technology*, pp. 14–26.

Stuart-Smith, J. *et al.* (2015) 'The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian', *Laboratory Phonology*, 6(3–4), pp. 505–549. doi: 10.1515/lp-2015-0015.

Stuart-Smith, J., Timmins, C. and Wrench, A. (2003) 'Sex and gender differences in Glaswegian /s/', *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 1851–1854. Available at: http://eresearch.qmu.ac.uk/2252/.

Sulpizio, S. *et al.* (2015) 'The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages', *PLoS ONE*, 10(7), pp. 1–38. doi: 10.1371/journal.pone.0128882.

Sulpizio, S. *et al.* (2019) 'Auditory Gaydar : Perception of Sexual Orientation Based on Female Voice', *Language and Speech*, pp. 1–23. doi: 10.1177/0023830919828201.

TheLanguageArchive (2018) 'ELAN'. Nijmegen: Max Planck Institute for Psycholinguistics.

Thomas, E. R. (2011) *Sociophonetics: An Introduction*. London: Palgrave Macmillan.

Thorpe, D. (2014) *Do I Sound Gay?* USA: IFC Films.

Tompkinson, J. and Watt, D. (2018) 'Assessing the abilities of phonetically untrained listeners to determine pitch and speaker accent in unfamiliar voices', *Language and Law / Linguagem e Direito*, 5(1), pp. 19–37.

Traunmüller, H. and Eriksson, A. (1995) 'The frequency range of the voice fundamental in the speech of male and female adults', *Consulté le*, 12(2), p. 2013.

Trudgill, P. (1974) *The social differentiation of English in Norwich*. Cambridge University Press.

Waksler, S. (2001) 'Pitch range and women's sexual orientation', *Word*, 52(1), pp. 69–77. doi: 10.1080/00437956.2001.11432508.

Wales, K. (2000) 'North and South: An English linguistic divide?', *English Today*, 16(1), pp. 4–15. doi: 10.1017/S0266078400011378.

Watt, D. (2013) 'Sociolinguistic variation in vowels', in Ball, M. and Gibbon, F. (eds) *Handbook of Vowels and Vowel Disorders*. New York: Taylor & Francis LLC, pp. 377–424.

Wells, J. C. (1986) *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Wells, J. C. (2006) *English Intonation: An Introduction*. Cambridge University Press.

Wiegman, R. and Wilson, E. A. (2015) 'Introduction: Antinormativity's queer conventions', *Differences*, 26(1), pp. 1–25. doi: 10.1215/10407391-2880582.

Wrench, A. A. (1995) 'Analysis of fricatives using multiple centres of gravity', *Proc. International Congress of Phonetic Sciences*, 4, pp. 460–463. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.5666&rep=rep1&type=

pdf.

Zimman, L. (2012) *Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people*, *Linguistics Graduate Theses & Dissertations*. Boulder: University of Colorado.

Zimman, L. (2017) 'Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants', *Linguistics*, 55(5), pp. 993–1019. doi: 10.1515/ling-2017-0018.

Zimman, L., Davis, J. L. and Raclaw, J. (eds) (2014) *Queer Excursions*. Oxford: Oxford University Press.

Zosky, D. L. and Alberts, R. (2016) 'What's in a name? Exploring use of the word queer as a term of identification within the college-aged LGBT community', *Journal of Human Behavior in the Social Environment*. Routledge, 26(7–8), pp. 597–607. doi: 10.1080/10911359.2016.1238803.

Appendix 1. Pilot Study Questionnaire

Title of project: Speech Variation within a Lesbian Community

Lead researcher: Salina Cuddy

Questionnaire

Age:

Gender:

Height: Weight (optional):

1.   What is your occupation?

2.   Where did you grow up?

3.   Where do you currently live? How long have you lived in this town/city?

4.   Your education: please circle one answer:

        I left school aged 14-16   I left school aged 17-19   I am at

        university

        I have completed a college or university degree

5.   How would you describe your sexual orientation?

6.   How would you describe your ethnicity?

Are you interested in having an article based on this research e-mailed to you?

If yes to the above question, please provide your e-mail:

Appendix 2. Short Story

Fern's Star Turn

Fern was a nurse from Harrogate who was always a happy-go-lucky person. One winter morning she was drawing a bath and washing her face with a cloth, when she saw a letter come through the door. She got a lot of letters, but when she went to look, this one caught her eye. It was from Paul, her father.

Paul was a keen dancer who had won many competitions in the past with his partner Pam. Their speciality was square dancing. In the letter, Paul explained that the International Square Dancing Championships were being held in New York City the following week, but unfortunately, Pam had just been admitted to hospital. She had managed to trap her foot in a bus door and had broken her leg and her nose when the bus moved off.

"Poor Pam", Fern thought, "what a daft thing to do! She won't be able to strut her stuff with Dad next week. They'll have no choice but to pull out, and it's too near to the competition to get the price of their tickets back."

Just then Fern had a fantastic idea and said, "I'm not half the dancer Pam is, but maybe I could stand in for her." Straight away she got on the phone to the travel agent and booked her flight, and then rang Paul to tell him what she had decided to do. She could tell her father was really happy. He suggested that they arrange to stay with friends on their farm outside the city, as the only hotels he could afford in New York looked rather seedy.

On Thursday the following week she got up at the crack of dawn to make a start on packing her kit for the trip. She knew that the north wind in

New York could be very cold in winter, so she grabbed her fleece jacket and her fur hat. She also packed the beautiful gold dress that Pam had made
for the competition, but it was quite bulky and she had to force her case closed by pressing down on the lid with the palm of each hand.

Paul and Fern drove south to the airport and shortly after checking in they boarded the plane. Their flight passed quickly and it seemed like no time before they were being greeted by Paul's friends Don and Sarah, who drove them to their pretty farmhouse surrounded by fir trees. On their farm there were horses neighing, sheep baa-ing, pot-bellied pigs, a pet goose called Rhonda and eight breeds of goat. That evening they were treated to a great feast of cured pork, which Sarah served out on large white plates. "Boy", thought Fern, "I'll need to watch my weight if I'm going to fit into Pam's gold dress."

The day of the competition it was pouring with rain, but Paul and Fern were too excited to care. They got dressed and made their way downstairs to Don's car. But disaster struck when the car wouldn't start. "What's wrong with it?" shouted Sarah from the house. "Have you got a toolkit in the boot?" Paul suggested to Don. "It's no good," Don sighed. "We'll have to call a cab, but it'll take a while to get to the city. It's a lot farther than you might think."

Fern and Paul made it to the competition with only seconds to spare.
They were out of breath and found it hard to remember the steps. However, they danced like champions and the judges were bowled over. They had no choice but to award them the first prize: a thousand dollars. Against the odds they had achieved their goal. Fern had made her pa a proud man. What a shame that half the prize money went on the taxi fare home!

Appendix 3. Final Questionnaire

Title of project: Speech Variation within a Football Community

Lead researcher: Salina Cuddy

Questionnaire

Age:

Gender:

Height:

7.   What is your occupation?

8.   Where did you grow up?

9.   Where do you currently live? How long have you lived in this town/city?

10.   Your education: please check one answer:

☐ I left school aged 14-16

☐ I left school aged 17-19

☐ I am at university

☐ I have completed a college or university degree

11.   How would you describe your sexual orientation?
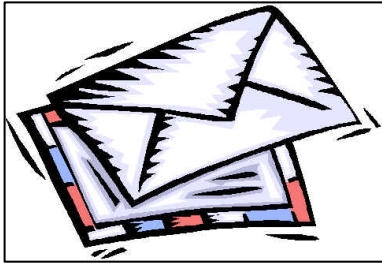
12.   How would you describe your ethnicity?

13.   How would you describe your social class?

14.   Are you or have you ever been a regular smoker?
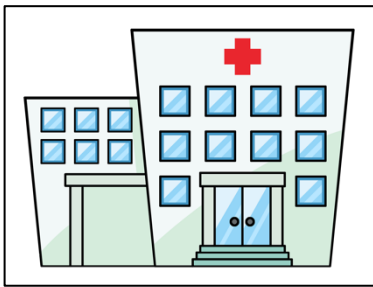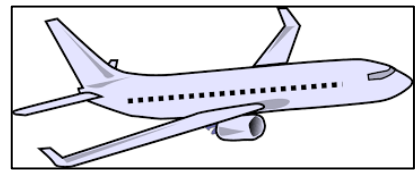
Appendix 4. Short Story Picture Guide

1.

2.

3.

4.

5.

6.

7.

8.

Appendix 5. Praat script: /s/ measurements

```
# SC 16/11/17, updated to get information without having
to delete anything from the file
# SC 31/8/17, updated for word
# SC 11/10/16, updated for surrounding sounds
# SC 13/7/16, updated for filter and only capturing /s/
# SC 7/6/16, based on
MeasureSpectralMomentsCreateTableExcel_Nat2.praat
# NF 14/02/11, based on ObsturentMeasures_Haynes.praat
and MeasureSpectralMoments_Hoole.praat




no_of_intervals = Get number of intervals... 1

clearinfo

full_name$ = selected$ ()
    name$ = extractLine$ (full_name$, " ")

# Create a table (named 'results') and define the column
labels
    Create Table... results 0 13
    Set column label (index)... 1 namefile
    Set column label (index)... 2 start
    Set column label (index)... 3 dur
    Set column label (index)... 4 centre
    Set column label (index)... 5 standdev
    Set column label (index)... 6 skewness
    Set column label (index)... 7 kurtosis
    Set column label (index)... 8 amp
    Set column label (index)... 9 peak
    Set column label (index)... 10 fricative
    Set column label (index)... 11 word
    Set column label (index)... 12 previous
    Set column label (index)... 13 following

select TextGrid 'name$'

j = 1

for i to no_of_intervals
    text$ = Get label of interval... 1 i
    if text$ = "S"

        i_min = i - 1
        i_plus = i + 1
```

```
        pre$ = Get label of interval... 1 i_min
        fol$ = Get label of interval... 1 i_plus
        start = Get starting point... 1 i
        end = Get end point... 1 i
        duration = end - start
        mid_point = start + duration / 2
        word_interval_no = Get interval at time... 2
mid_point
        word$ = Get label of interval... 2
word_interval_no


    # Get average amplitude of noise - normalisation??
        select Sound 'name$'
        To Intensity... 70 0 yes
        amplitude = Get mean... start end sones
        Remove


    # Create spectrum
        select Sound 'name$'

        Edit
        editor Sound 'name$'
        Select... start end
        View spectral slice
        Close
        endeditor
        Rename... 'name$'_slice


        select Spectrum 'name$'_slice
        #Cepstral smoothing... 500
        Filter (pass Hann band)... 1000 22050 100
        Rename... 'name$'_slice_smooth


    # Get peak (frequency of max. amplitude in Ltas)
        select Spectrum 'name$'_slice_smooth
        To Ltas (1-to-1)
        select Ltas 'name$'_slice_smooth
        peak = Get frequency of maximum... 0 0 Cubic
        Remove


    # Get spectral moments (power=2)
        select Spectrum 'name$'_slice_smooth
        centre = Get centre of gravity... 2
        sd = Get standard deviation... 2
```

```
            skewness = Get skewness... 2
            kurtosis = Get kurtosis... 2
            Remove


    # Get fricative
        select TextGrid 'name$'
        fric$ = Get label of interval... 1 i



    # Put everything in the table (named 'results')
        select Table results
        Append row
        Set string value... j namefile 'name$'
        Set numeric value... j start 'start'
        Set numeric value... j dur 'duration'
        Set numeric value... j centre 'centre'
        Set numeric value... j standdev 'sd'
        Set numeric value... j skewness 'skewness'
        Set numeric value... j kurtosis 'kurtosis'
        Set numeric value... j amp 'amplitude'
        Set numeric value... j peak 'peak'
        Set string value... j fricative 'fric$'
        Set string value... j word 'word$'
        Set string value... j previous 'pre$'
        Set string value... j following 'fol$'
        j = j + 1

    select TextGrid 'name$'

    endif
endfor

# Put everything in an text file
    select Table results
     Write to table file... 'name$'_results.txt
```

Appendix 6. IViE Corpus Sentences

1. We live in Ealing.
2. You remembered the lillies.
3. We arrived in a limo.
4. They are on the railings.
5. We were in yellow.
6. He is on the lilo.
7. You are feeling mellow.
8. We were lying.
9. He is on the lilo?
10. You remembered the lillies?
11. You live in Ealing?
12. May I lean on the railings?
13. May I leave the meal early?
14. Will you live in Ealing?
15. Where is the manual?
16. When will you be in Ealing?
17. Why are we in a limo?
18. Are you growing limes or lemons?
19. Is his name Miller or Mailer?
20. Did you say mellow or yellow?
21. Do you live in Ealing or Reading?
22. Did he say lino or lilo?

Appendix 7. Praat Script: F0 manipulation

```
## shift pitch up or down

# choose directory where wav files are saved
directory$ = "/Users/salinauk/Desktop/Tester/"

# create string list for all wav files in the directory
Create Strings as file list... list 'directory$'*.wav
number_of_files = Get number of strings

# start loop
for x from 1 to number_of_files
        select Strings list
        current_file$ = Get string... x

# select sound file
Read from file... 'directory$''current_file$'

# create pitch object
do ("To Manipulation...", 0.01, 75, 300)

# open manipulation editor, shift pitch, save new sound
to object list, close editor
    select all
    fileName$ = selected$("Manipulation", x)
    select Manipulation 'fileName$'
    Edit
    editor Manipulation 'fileName$'
    Select... 0 20
        do ("Shift pitch frequencies...", -20, "Hertz")
        do ("Publish resynthesis")
    Close
    endeditor
        Rename... 'fileName$'_DOWN20

# save new sound
        select Sound 'fileName$'_DOWN20
        Save as WAV file...
/Users/salinauk/Desktop/Tester/Down
20/'fileName$'_DOWN20.wav

endfor
```